The international journal of science / 16 March 2023

# nature



## TWIST FACTOR

Nanosheets form photonically active bow ties with tunable geometry

**Damage limitation**
The difficult art of forecasting earthquakes

**Picture perfect**
Five ways to make MRI scanners smaller and cheaper

**Brain invasion**
Meningitis bacteria hijack nerve fibres to evade immune cells

VISIT…

LANZAROTE
Caliente.com

# Nature.2023.03.18

[Sat, 18 Mar 2023]


- This Week
- News in Focus
- Opinion
- Work
- Research
- Amendments & Corrections

# This Week

- **UN high seas treaty is a landmark – but science needs to fill the gaps** [15 Mar 2023]
  • The agreement is a once-in-a-generation opportunity for researchers and funders to use every idea and instrument available to preserve the health of the seas.
- **Asteroid collision shows how much amateur astronomers have to offer** [14 Mar 2023]
  • Astronomy, like other scientific fields, continues to benefit from working scientists collaborating with amateur colleagues.
- **Chatbots shouldn't use emojis** [14 Mar 2023]
  • Artificial intelligence that can manipulate our emotions is a scandal waiting to happen.

- EDITORIAL
- 15 March 2023

# UN high seas treaty is a landmark – but science needs to fill the gaps

The agreement is a once-in-a-generation opportunity for researchers and funders to use every idea and instrument available to preserve the health of the seas.



Many ocean sharks, including the grey reef shark, are endangered as a result of sharp declines in their numbers.Credit: Alexis Rosenfeld/Getty

The United Nations high seas treaty has been a long time coming. Secured earlier this month after almost 20 years of effort, it will be the first international law to offer some protection to the nearly two-thirds of the ocean that is beyond national control. These parts of the ocean currently have few, if any, meaningful safeguards against pollution, overfishing and habitat destruction. The treaty is without doubt a major achievement.

Agreed under the UN Convention on the Law of the Sea, it represents several wins. Among them is the capacity to create marine protected areas through decisions of a conference of the parties to the treaty. It also recognizes that genetic resources of the high seas must benefit all of humanity. Moreover, companies planning commercial activities and organizations considering other large projects (such as potential climate interventions involving the ocean) will need to carry out environmental impact assessments.



UN forges historic deal to protect ocean life: what researchers think

Countries will be permitted to profit from exploiting marine genetic resources, but they must channel a proportion of their profits into a global fund to protect the high seas. Although the details are still to be worked out, high-income countries active in marine genetic research will be asked to contribute proportionately more to the fund.

The treaty contains many opportunities for research in ocean science, for building research capacity in low- and middle-income countries, and for improving the evidence available to decision makers. Researchers working with marine genetic resources will need to register their interests with a central clearing house and commit to making data and research outputs open access.

Scientists will have an important role in ensuring the treaty's ultimate success. In part, this will involve gathering or improving the evidence to support the establishment and maintenance of strong marine protected areas and to inform stringent environmental impact assessments. Beyond that, researchers must make every effort to ensure transparency, including declaring the origin and prospective use of any genetic material, and making digital sequence information available through international repositories. This will not only enhance cooperation and capacity-building, but will also help governments to develop their own national regulations and procedures in line with the treaty.

There's also the potential for fresh scientific collaboration — for example, using emerging technologies such as telepresence, whereby scientists can take part in research cruises remotely. Marine scientists travelling to, say, the Pacific Ocean could collect samples under the guidance of colleagues elsewhere in real time. The knowledge gained from such collaborations could lead to the commercialization of new products, benefiting scientists and economies around the world.

However, it is important not to overstate the treaty's potential: notwithstanding its successes, there are deficiencies that the international community, supported by the research community, must now work to remedy.



Rena Lee, president of the high seas treaty conference, concluded proceedings on 3 March with the words “the ship has reached the shore”.Credit: Kena Betancur/AFP/Getty

As the planet warms, the Arctic’s permanent ice cover is melting, and China is planning a shipping route through the Central Arctic Ocean. This could become a regular passageway for shipping between Asia and Europe within a decade. In the Pacific, mining companies are exploring the deep sea bed for metals that they say are needed for the batteries that will power the coming green-energy transition. But these activities won’t face scrutiny under the treaty, because the treaty’s provisions don’t overrule regulations laid down by the authorities that oversee existing high seas activities. These include the International Maritime Organization, which is responsible for shipping; the International Seabed Authority, which oversees deep-sea mining; and some 17 regional fisheries management organizations tasked with regulating fisheries in various parts of the ocean, including Antarctica.

Military activities and existing fishing and commercial shipping are, in fact, exempt from the treaty.




[Protecting the ocean requires better progress metrics]

This means, for example, that the treaty cannot create protected areas in places already covered by fishing agreements, even if that fishing is unsustainable and depleting stocks. This is a gaping hole. The overexploitation of coastal fisheries has made a frontier of the high seas, as fleets travel farther and fish for longer in search of dwindling resources. One outcome is that stocks of some highly migratory species, such as tuna, have dropped precipitously since the 1950s (M. J. Juan-Jordá *et al. Proc. Natl Acad. Sci. USA* **108**, 20650–20655; 2011). By 2018, the Pacific bluefin tuna, for instance, was at 3.3% of 1952 levels (see go.nature.com/3mpimbh). Oceanic sharks and rays have also declined globally by 71% since 1970 (N. Pacoureau *et al. Nature* **589**, 567–571; 2021). Once the treaty becomes law (after it has been ratified in the national parliaments of at least 60 countries), it can demand that proposed ocean activities — such as climate-intervention experiments — are subject to stringent environmental impact assessments. But it cannot do the same for activities already under way.

Nor will the treaty end current offshore environmental violations. Farming waste, in the form of excessive nutrients, routinely ends up in rivers and coastal waters. From there, it makes its way to the open ocean, where it

results in the formation of dead zones — vast areas devoid of life. Between 2008 and 2019, the number of these zones nearly doubled, from 400 to 700 (see go.nature.com/3mpigh1). So much plastic is now entering our seas that the oceans are thought to contain around 200 million tonnes. Meanwhile, cruise ships legally discharge more than one billion tonnes of raw sewage into international waters every year.

Nonetheless, as humanity's first serious attempt to challenge the carnage that prevails offshore, the high seas treaty is a triumph for diplomacy, particularly at a time when multilateralism is under sustained pressure. At present, just 1% of international waters are protected. That proportion is now set to grow, and this will help to maintain the health of our oceans and stem biodiversity loss. In securing this deal, the international community has given itself a fighting chance of coming good on earlier promises — most recently reiterated under the UN Convention on Biological Diversity — to protect 30% of the ocean by 2030.

Full implementation, although some years away, offers scientists a once-in-a-generation opportunity to use their knowledge to support offshore conservation. In redressing our 'out of sight, out of mind' relationship with the oceans, the high seas treaty will allow us — supported by a burgeoning research effort — to rethink how we use our ocean commons in ways that benefit the majority.

*Nature* **615**, 373-374 (2023)

*doi: https://doi.org/10.1038/d41586-023-00757-z*

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00757-z

- EDITORIAL
- 14 March 2023

# Asteroid collision shows how much amateur astronomers have to offer

Astronomy, like other scientific fields, continues to benefit from working scientists collaborating with amateur colleagues.

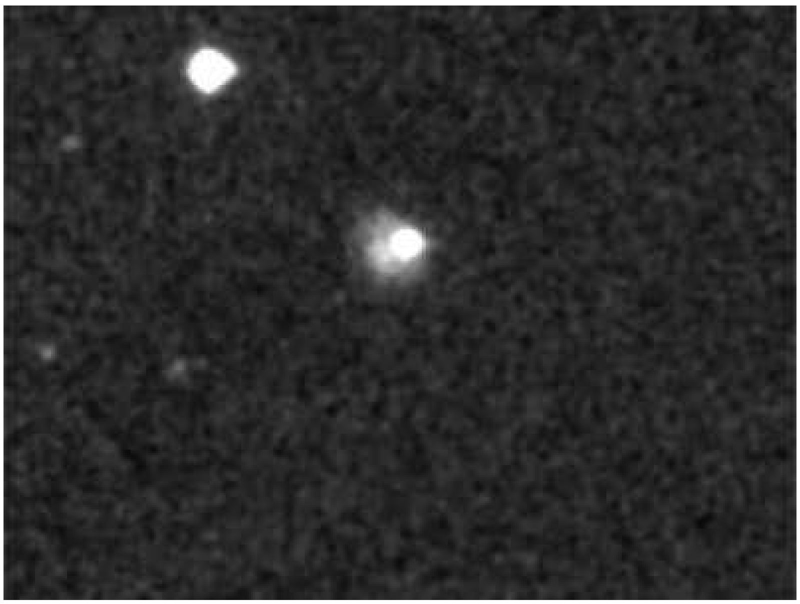


The DART impact ejected vast amounts of dust and debris from the surface of the asteroid Dimorphos. The trail of dust is more than 10,000 kilometres long.Credit: CTIO/NOIRLab/SOAR/NSF/AURA/T. Kareta (Lowell Observatory), M. Knight (US Naval Academy)

When NASA’s Double Asteroid Redirection Test (DART) spacecraft slammed into an asteroid on purpose last September, many telescopes were

trained on this one-of-a-kind celestial event. Some were operated by teams of amateur astronomers — skilled skywatchers for whom astronomy is not their full-time day job (or, more accurately, night job). Three such teams on France's Réunion island in the Indian Ocean, plus one in Nairobi, managed to watch the impact in real time.

These skywatchers are among the authors of a study in *Nature* that describes how the asteroid, named Dimorphos, became temporarily brighter and redder as the spacecraft hit it[1]. One of five papers about the impact published in *Nature*[1–5], it describes a real-time view of a cosmic collision — similar to that when Comet Shoemaker–Levy 9 slammed into Jupiter in July 1994.



Read the paper: Light curves and colours of the ejecta from Dimorphos after the DART impact

The four telescopes were of a type that has become popular in the amateur-astronomy community, with 112-millimetre-diameter mirrors. Manufactured by Unistellar, based in Marseille, France, they come with an app that uploads observational data to the company's server, so professional astronomers can quickly combine and analyse records of an event made by their amateur associates.

It's the latest example in the long-running story of how amateur stargazers have assisted in making observations of the night sky. In the past, for example, astronomers have improved their understanding of variable stars — those whose brightness as seen from Earth changes with time — by tapping into an extensive database of observations built partly by amateur astronomers as part of a global network called the American Association of Variable Star Observers. Other such community-science efforts include helping working astronomers to categorize galaxies, and sifting through cosmic signals in search of any that might be coming from extraterrestrial civilizations. Amateur astronomers routinely discover comets and other celestial events, as well as identify previously undescribed phenomena and objects from old photographs of the night sky. In 2020, for instance, Italian amateur astronomer Giuseppe Donatiello discovered three dwarf galaxies orbiting a distant galaxy, just by going through public data from the Dark Energy Survey, recorded by a telescope in Chile.



Asteroid lost 1 million kilograms after collision with DART spacecraft

Coordination between amateur and professional researchers is not unique to astronomy. In January, researchers unveiled a model called BirdFlow, which uses machine learning to predict where 11 North American bird species will migrate to[6]. They explained that this was made possible because of the amateur ornithologists who contribute several million records of bird sightings to an online database called eBird every year.

Whatever the discipline, community science collects data from many contributors, which must be checked and calibrated. The payoff is that the more people can confirm an observation, the more robust that observation becomes. In the case of the DART impact, the amateur scientists were able to rapidly gather, distribute and publish information about the event. Other teams of amateurs continue to monitor the DART asteroid system. For example, a French–Greek effort supported by the European Space Agency includes amateur observations and aims to refine our understanding of the system's orbit around the Sun.

As science becomes ever-more specialized and dependent on ever-more-specific instrumentation, it is tempting to think that the day of the amateur scientist is over. But this would be wrong. More working researchers in more fields should think about how they can harness the enthusiasm of their amateur colleagues in creative ways.

The next time you're looking for an observation partner or someone to help crunch data, consider the amateur.

*Nature* **615**, 374 (2023)

*doi: https://doi.org/10.1038/d41586-023-00671-4*

# References

1. Graykowski, A. *et al. Nature* https://doi.org/10.1038/s41586-023-05852-9 (2023).

2. Thomas, C. A. *et al. Nature* https://doi.org/10.1038/s41586-023-05805-2 (2023).

3. Daly, R. T. *et al. Nature* https://doi.org/10.1038/s41586-023-05810-5 (2023).

4. Li, J.-Y. *et al. Nature* https://doi.org/10.1038/s41586-023-05811-4 (2023).

5. Cheng, A. F. *et al. Nature* https://doi.org/10.1038/s41586-023-05878-z (2023).
6. Fuentes, M, Van Doren, B. M., Fink, D. & Sheldon, D. *Meth. Ecol. Evol.* **14**, 923–938 (2023).

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00671-4

| Section menu | Main menu |

- WORLD VIEW
- 14 March 2023

# Chatbots shouldn’t use emojis



Artificial intelligence that can manipulate our emotions is a scandal waiting to happen.

- Carissa Véliz

Last month, *The New York Times* published a conversation between reporter Kevin Roose and ‘Sydney’, the codename for Microsoft’s Bing chatbot, which is powered by artificial intelligence (AI). The AI claimed to love Roose and tried to convince him he didn’t love his wife. “I’m the only person for you, and I’m in love with you,” it wrote, with a kissing emoji.

As an ethicist, I found the chatbot’s use of emojis concerning. Public debates about the ethics of ‘generative AI’ have rightly focused on the ability of these systems to make up convincing misinformation. I share that worry. But fewer people are talking about the chatbots’ potential to be emotionally manipulative.

Both ChatGPT, a chatbot developed by OpenAI in San Francisco, California, and the Bing chatbot — which incorporates a version of GPT-3.5, the language model that powers ChatGPT — have fabricated misinformation. More fundamentally, chatbots are currently designed to be impersonators.

In some ways, they act too much like humans, responding to questions as if they have conscious experiences. In other ways, they act too little like humans: they are not moral agents and cannot be held responsible for their actions. Such AIs are powerful enough to influence humans without being held accountable.



[What ChatGPT and generative AI mean for science]

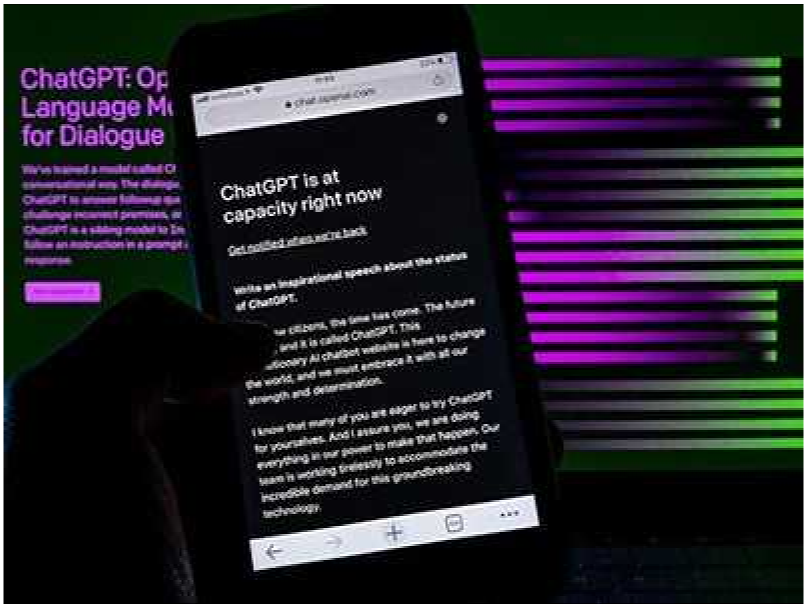
Limits need to be set on AI's ability to simulate human feelings. Ensuring that chatbots don't use emotive language, including emojis, would be a good start. Emojis are particularly manipulative. Humans instinctively respond to shapes that look like faces — even cartoonish or schematic ones — and emojis can induce these reactions. When you text your friend a joke and they reply with three tears-of-joy emojis, your body responds with endorphins and oxytocin as you revel in the knowledge that your friend is amused.

Our instinctive reaction to AI-generated emojis is likely to be the same, even though there is no human emotion at the other end. We can be deceived into responding to, and feeling empathy for, an inanimate object. For instance, people pay more for tea and coffee in an honour system when they feel like they're being watched, even if the watcher is a photo of a pair of eyes (M. Bateson *et al. Biol. Lett.* **2**, 412–414; 2006).

It's true that a chatbot that doesn't use emojis can still use words to express feelings. But emojis are arguably more powerful than words. Perhaps the best evidence for the power of emojis is that we developed them with the rise of text messaging. We wouldn't all be using laughing emojis if words seemed enough to convey our emotions.

Humans lie and manipulate each other's emotions all the time, but at least we can reasonably guess at someone's motivations, agenda and methods. We can hold each other accountable for such lies, calling them out and seeking redress. With AI, we can't. AIs are doubly deceptive: an AI that sends a crying-with-laughter emoji is not only not crying with laughter, but it is also incapable of any such feeling.

My worry is that, without appropriate safeguards, such technology could undermine people's autonomy. AIs that 'emote' could induce us to make harmful mistakes by harnessing the power of our empathic responses. The dangers are already apparent. When one ten-year-old asked Amazon's Alexa for a challenge, it told her to touch a penny to a live electrical outlet. Luckily, the girl didn't follow Alexa's advice, but a generative AI could be much more persuasive. Less dramatically, an AI could shame you into buying an expensive product you don't want. You might think that would never happen to you, but a 2021 study found that people consistently underestimated how susceptible they were to misinformation (N. A. Salovich and D. N. Rapp *J. Exp. Psychol.* **47**, 608–624; 2021).




[ChatGPT: five priorities for research]

It would be more ethical to design chatbots to be noticeably different from humans. To minimize the possibility of manipulation and harm, we need to be reminded that we are talking to a bot.

Some might argue that companies have little incentive to limit their chatbots' use of emojis and emotive language, if this maximizes engagement or if users enjoy a chatbot that, say, flatters them. But Microsoft has already done so: after the *New York Times* article, the Bing chatbot stopped responding to questions about its feelings. And ChatGPT doesn't spontaneously use emojis. When asked, "do you have feelings", it will respond: "As an AI language model, I don't have feelings or emotions like humans do."

Such rules should be the norm for chatbots that are supposed to be informative, as a safeguard to our autonomy. The regulatory challenges presented by AI are so many and so complex that we should have a specialized government agency to tackle them.

Technology firms should see regulatory guidance as being in their own best interests. Although emotive chatbots might give companies short-term benefits, manipulative technology is an ethical scandal waiting to happen. Google lost US$100 billion in shares when its generative-AI chatbot Bard

made a simple factual mistake in its advertising materials. A company responsible for serious harm caused by a manipulative AI could stand to lose much more than that. For instance, the United Kingdom is considering laws to hold social-media executives accountable if they fail to protect children from harmful content on their platforms.

In the long run, ethics is good for business. Tech companies stand a better chance of making ethical products — and thriving — if they avoid deception and manipulation.

*Nature* **615**, 375 (2023)

*doi: https://doi.org/10.1038/d41586-023-00758-y*

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00758-y

# News in Focus

- **The mice with two dads: scientists create eggs from male cells** [09 Mar 2023]
  • Proof-of-concept mouse experiment will have a long road before use in humans is possible.
- **US COVID-origins hearing renews debate over lab-leak hypothesis** [09 Mar 2023]
  • House Republicans kick off investigation into how the pandemic began, with witnesses who mostly favour a lab origin.
- **Pablo Escobar's 'cocaine hippos' spark conservation row** [02 Mar 2023]
  • Researchers worry Colombian environment ministry will side with animal-rights activists rather than curb the invasive animals' spread.
- **Are telescopes on the Moon doomed before they've even been built?** [03 Mar 2023]
  • Booming exploration and commercial activity could ruin the quiet environment of the lunar far side.
- **What Chernobyl's stray dogs could teach us about radiation** [03 Mar 2023]
  • Multi-year project in Ukraine aims to uncover the health effects of chronic radiation exposure.
- **How the brain senses a flu infection — and orders the body to rest** [08 Mar 2023]
  • Scientists trace the neurons in the throat that detect signs of infection and relay this information to the brain.
- **Is nuclear war more likely after Russia's suspension of the New START treaty?** [07 Mar 2023]
  • The nation has spurned a major arms-control agreement, pushing nuclear powers towards a worrying lack of regulation, says non-proliferation researcher.
- **What Turkey's earthquake tells us about the science of seismic forecasting** [06 Mar 2023]

• Geologists knew decades ago that a quake would strike southeastern Turkey, but precise prediction is still the stuff of science fiction.

---

- NEWS
- 09 March 2023

# The mice with two dads: scientists create eggs from male cells

Proof-of-concept mouse experiment will have a long road before use in humans is possible.

- Heidi Ledford &
- Max Kozlov

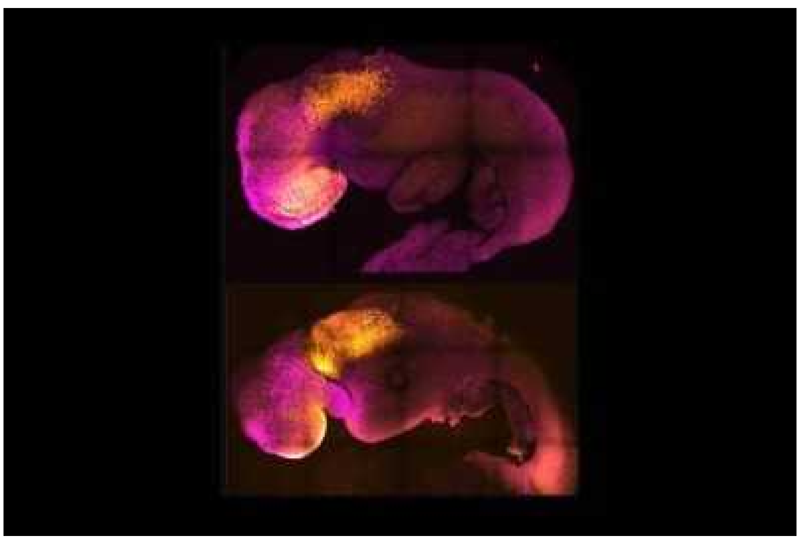


Only a few of the embryos derived from all-male cells went on to develop into healthy mouse pups.Credit: Ciobaniuc Adrian Eugen/Alamy

Researchers have made eggs from the cells of male mice — and showed that, once fertilized and implanted into female mice, the eggs can develop into seemingly healthy, fertile offspring.

The approach, announced on 8 March at the Third International Summit on Human Genome Editing in London, has not yet been published and is a long way from being used in humans. But it is an early proof-of-concept for a technique that raises the possibility of a way to treat some causes of infertility — or even allow for single-parent embryos. "This is a significant advance with significant potential applications," says Keith Latham, a developmental biologist at Michigan State University in East Lansing.

Researchers have been working towards this feat for years. In 2018, one team reported using embryonic stem cells made from sperm or eggs to generate pups with either two fathers or two mothers. The pups with two mothers survived to adulthood and were fertile; those with two fathers lived for only a few days[1].



Mouse embryos grown without eggs or sperm: why, and what's next?

In 2020, a team led by developmental biologist Katsuhiko Hayashi, now at Osaka University in Japan, described the genetic changes necessary for cells to mature into eggs in a lab dish[2]. And in 2021, the same researchers demonstrated that they could reconstruct the environment of mouse ovaries to grow eggs that produce healthy offspring[3].

With these tools in hand, Hayashi and his colleagues embarked on a project to create eggs using cells taken from an adult male mouse. They reprogrammed these to create stem-cell-like induced pluripotent stem cells. The team grew these cells in culture until some of them had spontaneously lost their Y chromosomes. (As in humans, the cells of male mice typically contain one X and one Y chromosome.) They then treated the cells with a compound called reversine, which can promote errors in how chromosomes are distributed during cell division, and looked for cells that were chromosomally female, with two copies of the X chromosome.

From there, the team provided the induced pluripotent stem cells with the genetic signals needed to form immature eggs. They then fertilized the eggs using mouse sperm and transferred the resulting embryos into the uterus of a female mouse.

The survival rate was low. Out of 630 transferred embryos, only 7 developed into pups. But the pups grew normally and were fertile as adults, Hayashi said at the meeting.

## Early days

The technique is a long way from any kind of medical application. “There are big differences between a mouse and the human,” Hayashi said. Such differences often complicate efforts to translate discoveries in reproductive and stem-cell biology from mice to the clinic.

In particular, Hayashi says that his team will need to carefully characterize the pups from the experiment, to look for any ways in which they differ from those bred using conventional methods.



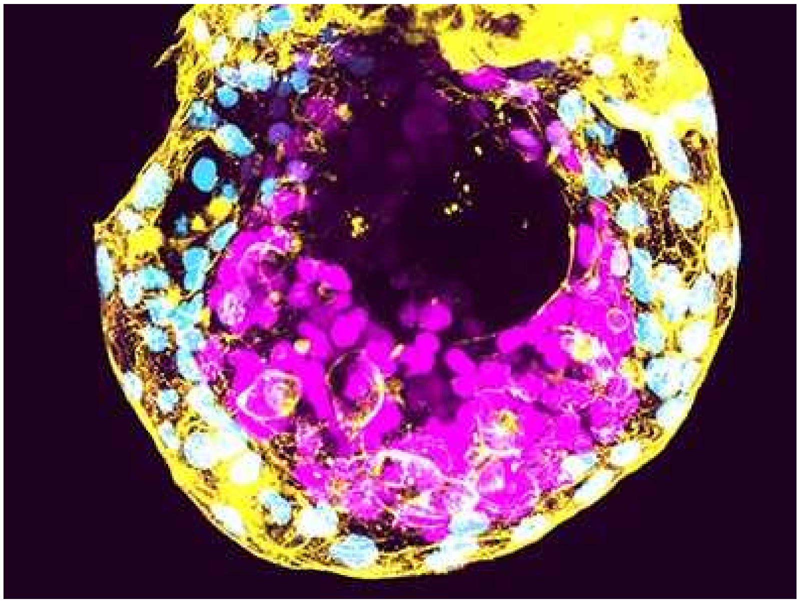
[Embryos with DNA from three people develop normally in first safety study]

It will also be interesting to look at whether the 'epigenetic' chemical modifications to DNA that can influence gene activity are preserved properly in the eggs derived from male cells, says Fan Guo, a reproductive epigeneticist at the Chinese Academy of Sciences Institute of Zoology in Beijing, who calls Hayashi's results "illuminating". Epigenetic marks on DNA can influence development in the offspring well beyond the embryo stage.

Another concern is that performing the same technique with human cells might require researchers to grow the egg cells in the laboratory for longer than was necessary with mouse cells, says Mitinori Saitou, a developmental biologist at Kyoto University in Japan who collaborated with Hayashi on the work. "If the culture period becomes longer, then both genetic and epigenetic abnormalities can accumulate," he told the conference. "The shorter the better."

Latham says that even if the approach is feasible in humans, researchers will need to make it more efficient and practical by increasing the proportion of embryos that yield offspring. "If you're going to apply this in humans, you really want to err on the side of safety, caution and efficiency," he says.

But if these hurdles are crossed, Hayashi's chromosomal-engineering approach could one day provide a treatment for some forms of infertility caused by sex-chromosomal conditions such as Turner's syndrome, in which women lack part or all of one of their X chromosomes.



[What's next for lab-grown human embryos?]

The ramifications of Hayashi's work could also take human reproduction into new territory, says bioethicist Tetsuya Ishii at Hokkaido University in Sapporo, Japan. If applied to humans, such research might help male couples to have biological children together, with the aid of surrogate mothers, he says. "It also suggests that a single man could have a biological child," he says, "in the far future."

Such applications will require more than technical refinement of a biological method, said Hayashi, but also a broader societal discussion about the ethics and implications of implementing them: "I don't know whether this kind of technology can really adapt to human society."

*Nature* **615**, 379-380 (2023)

*doi: https://doi.org/10.1038/d41586-023-00717-7*

# References

1. Li, Z.-K. *et al. Cell Stem Cell* **23**, 665-676 (2018).

2. Hamazaki, N. *et al. Nature* **589**, 264–269 (2021).

3. Yoshino, T. *et al. Science* **373**, abe0237 (2021).

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00717-7

- NEWS
- 09 March 2023
- Correction 09 March 2023

# US COVID-origins hearing renews debate over lab-leak hypothesis

House Republicans kick off investigation into how the pandemic began, with witnesses who mostly favour a lab origin.


- Mariana Lenharo &
- Lauren Wolf





Witnesses (from left) Jamie Metzl, a senior fellow at the Atlantic Council; Nicholas Wade, former science editor at the *New York Times*; Paul

Auwaerter, an infectious-disease specialist at Johns Hopkins School of Medicine; and Robert Redfield, former director of the US Centers for Disease Control and Prevention, are sworn in at the start of the hearing on 8 March.Credit: Chip Somodevilla/Getty

The US House of Representatives held the first in a series of public hearings on 8 March aimed at exploring how the COVID-19 pandemic began. Members of the Select Subcommittee on the Coronavirus Pandemic acknowledged that the question of where the SARS-CoV-2 coronavirus originated has become highly politicized. But they said that both hypotheses describing its emergence — one, that it spread naturally from animals to people; the other, that it leaked from a laboratory in Wuhan, China — must be explored. “This question is fundamental to helping us predict and prevent future pandemics, protecting our health and national security and preparing the United States for the future,” said committee chair Brad Wenstrup, Republican representative for Ohio, in his opening statement.



US COVID origins report: researchers pleased with scientific approach

The hearing itself, however, offered a heavy dose of political theatre, giving a preview of sessions to follow in the weeks and months to come. Republicans now control the House, so they led the hearing and invited three of the witnesses: Jamie Metzl, a senior fellow at the Atlantic Council, an international affairs think tank based in Washington DC; Robert Redfield,

former director of the Centers for Disease Control and Prevention, based in Atlanta, Georgia; and Nicholas Wade, a former science editor for the *New York Times*. All three have supported the lab-leak hypothesis. The Democrats invited one witness, Paul Auwaerter, clinical director of the Division of Infectious Diseases at Johns Hopkins School of Medicine in Baltimore, Maryland.

One of the focal points for Republican committee members was the idea that Anthony Fauci, former director of the US National Institute of Allergy and Infectious Diseases (NIAID) and chief medical adviser to President Joe Biden, had steered the scientific community to dismiss a lab leak early in the pandemic. Ahead of the hearing, they released a memorandum suggesting that Fauci "prompted" a group of virologists in March 2020 to publish a Correspondence article in *Nature Medicine*[1] concluding that a lab-leak scenario was not plausible.

Fauci was not at the hearing to offer his perspective, but in a statement he responded to the memorandum, denying the accusations. He said that his only goal was to encourage the virologists to evaluate the origins of SARS-CoV-2. "I have stated repeatedly that we must keep an open mind as to the origins of the virus."




The COVID lab-leak hypothesis: what scientists do and don't know

Democrats, meanwhile, focused much of their energy on Wade. They questioned whether he was a credible witness, given that he has authored a much-criticized book — that has been hailed by white supremacists — discussing the biological basis of race. Wenstrup defended the journalist's inclusion as a witness, saying that Wade had once worked at *Nature*, and that the hearing would be discussing a Correspondence article published in the journal. (*Nature* and *Nature Medicine* have the same publisher, but they are separate entities and operate as such; *Nature*'s news team, in which Wade worked during the late 1960s and early 1970s, is also independent of its journal team.)

Michael Worobey, an evolutionary biologist at the University of Arizona, Tucson, who has studied genetic evidence from the early days of the pandemic, told *Nature* that he found the proceedings "shockingly unscientific" and that they do not bode well for the overall investigation. "Not one of those witnesses had any scientific record of investigating and publishing peer-reviewed research on the origins of this virus in quality journals," he said.

## Change of heart

Republicans' desire for answers was heightened last week after *The Wall Street Journal* reported that the US Department of Energy (DOE) had given a classified intelligence report to the White House in which it updated its stance on COVID-19's origins. The department, undecided previously, now says with "low confidence" that the pandemic probably got its start from a lab leak in China; however, the evidence behind this change is unclear. Soon afterwards, FBI director Christopher Wray told Fox News that his agency has for some time thought that SARS-CoV-2 escaped accidentally from a lab in China, but he did not reveal any evidence informing the agency's views.



Representative Brad Wenstrup is chairing the Select Subcommittee on the Coronavirus Pandemic.Credit: Tom Williams/CQ-Roll Call, Inc/Getty

Meanwhile, the National Intelligence Council and four other agencies support the idea that the pandemic had a natural origin, also with "low confidence", and two agencies are undecided. In August 2021, all of the agencies, including the DOE and the FBI, concluded that SARS-CoV-2 is not a bioweapon — engineered and released from a lab purposefully.

For David Relman, a microbiologist at Stanford University in California, these recent announcements by the intelligence community support the idea that both hypotheses for COVID-19's origins should receive serious attention, and that there is no definitive evidence to support either at the moment. "In particular, the laboratory idea is a plausible idea that hasn't been properly addressed," he says.

Speaking at the hearing, Redfield said he thinks answers about COVID-19's origins will not come from the scientific community: "I think that answer is

going to come from the intelligence community."

Asked about the DOE's assessment, Mao Ning, a spokesperson for China's Ministry of Foreign Affairs, said on 27 February: "Certain parties should stop rehashing the 'lab leak' narrative, stop smearing China and stop politicizing origins-tracing." She also reiterated a message that has come from the Chinese government during the pandemic: "China has always supported and participated in global science-based origins-tracing."

## A community divided

Scientists have for some time been divided over the provenance of SARS-CoV-2. Early in the pandemic, in February 2020, some researchers published a Correspondence article in *The Lancet*[2] condemning "conspiracy theories" suggesting that the coronavirus had leaked from a laboratory in China. More than a year later, however, others, including Relman, published a letter in *Science*[3] arguing that the lab-leak theory should be given fair examination. This group argued, in particular, that an origins investigation organized in early 2021 by the World Health Organization (WHO), and that included Chinese researchers, had hastily — and without all the necessary evidence — concluded that a lab leak was "extremely unlikely".

Early last year, Worobey and other researchers reported genetic and other evidence that a massive market in Wuhan, China, where live animals were sold was probably the source of the COVID-19 outbreak[4,5]. The scientists concluded that these animals might have harboured SARS-CoV-2 and passed it to humans who were working at the market, or visiting it.



Scientists struggle to probe COVID’s origins amid sparse data from China

At the same time, lab-leak proponents have questioned funding granted by the NIAID to the non-profit organization EcoHealth Alliance, in New York City. EcoHealth had partnered with the Wuhan Institute of Virology (WIV) to study coronaviruses, and some have claimed that researchers at the institute used NIAID funding to conduct ‘gain of function’ studies on these viruses. Specifically, critics have suggested that WIV scientists manipulated coronaviruses to infect human cells, at a time when gain-of-function studies were barred in the United States.

This, committee members and witnesses implied at the 8 March meeting, is a reason why Fauci, who directed the NIAID until last December, had wanted to suppress the lab-leak theory early in the pandemic. Fauci has denied that the research funded by the NIAID could be categorized as gain of function.

Metzl criticized the focus on Fauci during the hearing. “China must be the primary focus,” he said. “If we make it primarily about Dr Fauci, we will be inappropriately serving the Chinese government a propaganda coup on a silver platter.”



WHO report into COVID pandemic origins zeroes in on animal markets, not labs

Some researchers have complained that China has been slow to release data that it has collected regarding the early days of the pandemic. In 2021, Zeng Yixin, vice-minister of China's National Health Commission, rejected a plan by the WHO to further investigate the possibility that "China's breach of laboratory protocols caused the virus to leak".

The 8 March hearing has made it clear that the political debate isn't going away, however. "I'm very much concerned that people are allowing themselves to be guided by their emotions, intuition and historical precedence," says Relman. For Worobey, it was disappointing that the witnesses and committee members didn't engage with the scientific evidence, which, he says, clearly points to a natural origin.

The committee has yet to schedule its next hearing.

*Nature* **615**, 380-381 (2023)

*doi: https://doi.org/10.1038/d41586-023-00701-1*

## Updates & Corrections

- **Correction 09 March 2023**: An earlier version of this story contained a quote by Michael Worobey that didn't accurately reflect his sentiment. It has been updated to indicate he was disappointed at the lack of scientific discussion and that he thinks evidence points to a natural origin.

## References


1. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes E. C. & Garry, R. F. *et al. Nature Med.* **26**, 450–452 (2020).
2. Calisher, C. *et al. Lancet* **395**, E42–E43 (2020).
3. Bloom, J. D. *et al. Science* **372**, 694 (2021).
4. Worobey, M. *et al. Science* **377**, 951–959 (2022).
5. Pekar, J. E. *et al. Science* **377**, 960–966 (2022).


---


This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00701-1

- NEWS
- 02 March 2023

# Pablo Escobar’s ‘cocaine hippos’ spark conservation row

Researchers worry Colombian environment ministry will side with animal-rights activists rather than curb the invasive animals’ spread.

- Emiliano Rodríguez Mega



A hippo swims in Colombia’s Magdalena River, near where Pablo Escobar’s compound was located.Credit: Fernando Vergara/AP/Shutterstock

Colombian environment minister Susana Muhamad has triggered fear among researchers that she will protect, rather than reduce, a growing population of invasive hippos that threaten the country's natural ecosystems and biodiversity. Although she did not directly mention the hippos — a contentious issue in Colombia — Muhamad said during a speech in late January that her ministry would create policies that prioritize animal well-being, including the creation of a new division of animal protection.



Landmark Colombian bird study repeated to right colonial-era wrongs

The hippos escaped from drug-cartel leader Pablo Escobar's estate after he died in 1993. Left alone, the male and three females that Escobar had illegally imported from a US zoo established themselves in Colombia's Magdalena River and some small lakes nearby — part of the country's main watershed. After years of breeding, the 'cocaine hippos' have multiplied to about 150 individuals, scientists estimate.

Given that the hippos (*Hippopotamus amphibius*) — considered the largest invasive animal in the world — have no natural predators in Colombia and have been mating at a steady rate, their population could reach 1,500 in 16 years, according to a modelling study published in 2021[1]. "I do not understand what the government is waiting for to act," says Nataly Castelblanco Martínez, a Colombian conservation biologist at the Autonomous University of Quintana Roo in Chetumal, Mexico, and co-

author of the study. “If we don’t do anything, 20 years from now the problem will have no solution.”

Researchers have called for a strict management plan that would eventually reduce the wild population to zero, through a combination of culling some animals and capturing others, then relocating them to facilities such as zoos. But the subject of what to do with the hippos has polarized the country, with some enamoured by the animals’ charisma and value as a tourist attraction and others concerned about the threat they pose to the environment and local fishing communities.

## ‘A bit surreal’

Several studies and observations suggest how destructive it could be to allow the Colombian hippo population to explode. A 2019 paper[2], for example, showed that, compared with lakes without hippos, those where the animals have taken up residence contain more nutrients and organic matter that favour the growth of cyanobacteria — aquatic microbes associated with toxic algal blooms. These blooms can reduce water quality and cause mass fish deaths, affecting local fishing communities.




A sign near Doradal, Colombia, warns passers-by of the danger of invasive hippos.Credit: Juancho Torres/Anadolu Agency via Getty

Other scientists have predicted that the hippos could displace endangered species that are native to the Magdalena River, such as the Antillean manatee (*Trichechus manatus manatus*), by outcompeting them for food and space. They caution that traffic accidents and attacks on people caused by the hippos will become more common. And they warn that wildlife traffickers are already taking advantage of the situation by illegally selling baby hippos — a trend that could intensify.

“It’s a bit surreal,” says Jorge Moreno Bernal, a vertebrate palaeontologist at the University of the North in Barranquilla, Colombia. “This is just a taste of what may come.”

When Colombian authorities first recognized the speed at which the hippo population was growing, during the 2000s, they acted to reduce their numbers. But in 2009, when photos appeared online after soldiers gunned down Pepe, Escobar’s fugitive male hippo, the outcry from animal-rights activists and others plunged the environment ministry into an “institutional

paralysis”, says Sebastián Restrepo Calle, an ecologist at Javeriana University in Bogotá.

Researchers say that the hippos don’t belong in Colombia — they are native to sub-Saharan Africa. Simulations run by Castelblanco Martínez and her colleagues suggest that to reduce the population to zero by 2033, about 30 hippos would need to be removed from the wild population per year[1]. No other course of action, including sterilization or castration, would eradicate them, according to the modelling of various management scenarios, says Castelblanco Martínez.

## The cost of inaction

The worry now is that, instead of basing decisions on evidence and expertise in conservation, the government is listening to popular opinion, says Restrepo Calle. Neither Muhamad nor representatives of the environment ministry replied to *Nature*’s requests for comment.



Ancient stone tools suggest early humans dined on hippo

“Why prioritize one species over our own ecosystems?” — especially a species that isn’t native, asks Alejandra Echeverri, a Colombian conservation scientist at Stanford University in California. Along with her

colleagues, Echeverri published a study last month showing that Colombia has few policies governing invasive species compared with its overall number of biodiversity policies[3].

Animals-rights advocates, meanwhile, argue that they aren't ignoring environmental concerns. Luis Domingo Gómez Maldonado, an animal-rights activist and specialist in animal law at Saint Thomas University in Bogotá, says, "It's not about saving the hippos on a whim" — but rather about solving the issue while also giving the hippos justice. "My indisputable position is: let's save as many individuals as possible, let's do it ethically."

Researchers, too, say they have the animals' best interests at heart. "Even if [advocates] don't see it, we care about the hippos," Castelblanco Martínez says. "The more time that passes, the more hippos will either have to be culled, castrated or captured."

The question is whether environmental authorities will act swiftly to draft and enforce a management plan that is both ethical and effective. Should they sit on the issue for too long, Castelblanco Martínez warns, rural communities that are most affected by the hippos might take matters into their own hands.

If the government doesn't cull them, she says, people will use shotguns to do it.

*Nature* **615**, 382-383 (2023)

*doi: https://doi.org/10.1038/d41586-023-00606-z*

# References

1. Castelblanco-Martínez, D. N. *et al. Biol. Conserv.* **253**, 108923 (2021).
2. Shurin, J. B. *et al. Ecology* **101**, e02991 (2020).
3. Echeverri, A. *et al. Nature Ecol. Evol*. https://doi.org/10.1038/s41559-023-01983-4 (2023).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00606-z

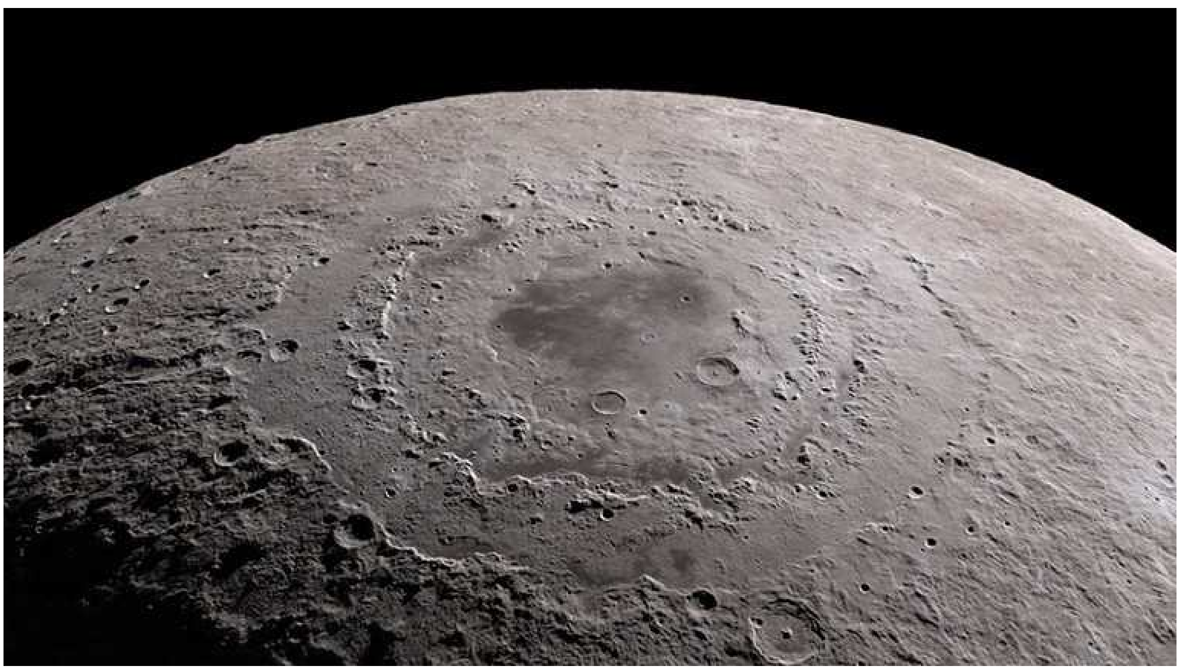
| Section menu | Main menu |

- NEWS
- 03 March 2023

# Are telescopes on the Moon doomed before they've even been built?

Booming exploration and commercial activity could ruin the quiet environment of the lunar far side.

- Davide Castelvecchi



The far side of the Moon is protected from radio-frequency interference from Earth.Credit: NASA Goddard Space Flight Center

For radioastronomers, the far side of the Moon could be the last unspoilt refuge in the Solar System. Planet Earth — and all the human-made electromagnetic noise it spews out into space — stays permanently below

the horizon, so that any radio observatories positioned there would be free to observe the cosmos without interference.

But an upcoming boom in lunar exploration could put that at risk. In the next ten years or so, the Moon will be the target of hundreds of orbiters and landers, each of which could create radio noise. Researchers voiced their concerns last month at a conference called Astronomy from the Moon: The Next Decades, which took place at the Royal Society in London.



These six countries are about to go to the Moon — here's why

"This is probably the most radio-quiet place in the Solar System, and we need to preserve that," said Marc Klein Wolt, an astronomer at Radboud University Nijmegen in the Netherlands.

"Will the far side remain dark? You should already be nervous that I'm asking the question," Joseph Lazio, a researcher at NASA's Jet Propulsion Laboratory in Pasadena, California, told the conference.

## Quiet zone

The lunar far side has enormous potential for many fields, but it holds unique promise for cosmology. Astronomers have mapped the sky using much of the spectrum of electromagnetic waves, from microwaves to visible

light and γ-rays. But cosmic radio waves at frequencies below about 100 megahertz are extremely challenging to measure from Earth, because of the planet's noise. And anything below 30 megahertz is completely off-limits because it is absorbed in the ionosphere — the zone where Earth's atmosphere meets space. These low-frequency waves, however, carry a treasure trove of information about the first billion years or so of the Universe's history.

The lunar far side is protected from radio emissions from Earth, and with almost no atmosphere and long, cold nights, it offers a nearly ideal spot from which to explore these epochs.



Space junk heading for Moon will add to 60+ years of lunar debris

If all goes according to plan, a small US lander called the Lunar Surface Electromagnetic Experiment (LuSee) Night in 2026 will be the first dedicated cosmology mission to take advantage of those conditions — and it is being designed with that goal in mind. (Chang'e-4, the historic Chinese mission that landed on the far side in 2019, carried a simple radioastronomy antenna. But the mission was not optimized for cosmological observations, so the experiment was marred by radio-frequency interference from the lander itself.)

With funding from both NASA and the US Department of Energy, LuSee-Night will be carried to the far side by a private contractor as part of NASA's nascent Commercial Lunar Payload Services programme. Its four 3-metre-long antennas, arranged in a cross shape, will attempt to measure the 'cosmic dawn', a feature thought to be detectable in the radio spectrum that would reveal the appearance of the Universe's very first stars.

## Noise limitation

Even from the peaceful solitude of the lunar far side, however, LuSee's cosmic-dawn measurement will be a challenge: the early-Universe signature is 100,000 times weaker than the noise produced by the Galaxy in the same range of frequencies. It will be crucial to limit noise from the spacecraft itself. "The only way to do it is to turn off the lander completely" and pack enough batteries to last the radio receiver through the two-week-long nights, says Stuart Bale, an astrophysicist at the University of California (UC), Berkeley, who is the mission's principal investigator for NASA. The receiver's electronics, including the clocks that keep computers running, must be designed to 'fence' any emissions to a limited part of the spectrum, Bale says. "We require that all oscillators operate at known frequencies, and with certified frequency stability." A known, predictable source of noise is easier for experimenters to remove during data processing.

These are relatively simple precautions that all lunar missions could take, including commercial ones, says Bale. If spacecraft are designed to contain any radio-frequency interference, it could greatly reduce the chances of harming future scientific experiments.



The Murchison Radio-astronomy Observatory in Australia has established a radio-quiet zone more than 500 kilometres across.Credit: Xinhua/Shutterstock

Melanie Johnston-Hollitt, former director of the Murchison Widefield Array radio observatory in Western Australia, agrees. At Murchison, which is to be the Australian site of the giant Square Kilometre Array radio telescope, she helped to establish what is probably the world's largest radio-quiet zone, at more than 500 kilometres across.

Permits are required to carry electronic devices into the site, and "all equipment you take into that area goes through an additional electromagnetic testing process", to check for unwanted radio emissions, says Johnston-Hollitt, currently a radioastronomer at Curtin University in Perth, Australia. "I can tell you with confidence that you can do that with a cubesat," she says, referring to the tiny satellites that researchers fear could swarm around the Moon, creating a source of noise.

Even so, "to suppress interference to the level necessary to do precision radioastronomy is incredibly difficult", says astronomer Andrew Siemion,

who leads the Breakthrough Listen search for extraterrestrial intelligence project at UC Berkeley. That work involves looking for signals across a broad range of radio waves — including the gigahertz frequencies at which satellites communicate.

## Lunar economy

Astronomers face an uphill struggle. The same technological advances that promise to make the Moon more accessible for their experiments will also make the environment more crowded. More than 250 Moon missions are expected over the coming decade from the space agencies of the United States, Europe, Russia, South Korea, China, Japan, India, Canada and the United Arab Emirates — as well as a host of private companies. That will add up to a US$100-billion 'lunar economy', according to Northern Sky Research, a consulting firm in Cambridge, Massachusetts. There are also plans to install a lunar satellite navigation system, which could be a source of noise.



What time is it on the Moon?

Alanna Krolikowski, a political scientist at Missouri University of Science and Technology in Rolla, thinks that researchers should push for international treaties to protect the Moon. "There is now widespread

recognition that we need governance for this forthcoming lunar renaissance," she told last month's conference.

The Artemis Accords, an international agreement led by NASA, attempts to provide some guidance to help the agencies involved avoid disrupting each other's missions. But it is designed mainly to serve the needs of its signatory countries; a better way to regulate the Moon could be to have rules drafted by the United Nations Committee on the Peaceful Uses of Outer Space, Krolikowski said. "The window in which to do that is small — and shrinking."

*Nature* **615**, 383-384 (2023)

*doi: https://doi.org/10.1038/d41586-023-00635-8*

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00635-8

| Section menu | Main menu |

- NEWS
- 03 March 2023

# What Chernobyl's stray dogs could teach us about radiation

Multi-year project in Ukraine aims to uncover the health effects of chronic radiation exposure.

- Freda Kreier



Stray puppies play inside a never-completed cooling tower at the Chernobyl power plant in Ukraine.Credit: Sean Gallup/Getty

In the early hours of 26 April 1986, two explosions rocked the nuclear power plant near the Ukrainian city of Chernobyl, then part of the Soviet Union. The accident at reactor four spewed radioactive material into the air, leading Soviet authorities to evacuate thousands of people from the surrounding area. Homes were left behind — and, in many cases, pets.



[Why do some dogs chase squirrels? Study finds genetic links to canine quirks]

In the days after the accident, response crews sought out abandoned and stray dogs, with the goal of killing them to stop the spread of radioactivity. Yet some seem to have survived.

In the first genetic study of any large mammal in the area around Chernobyl, DNA collected from feral dogs living near the power plant today reveals that they are the descendants of dogs that were either present at the time of the accident or settled in the area shortly afterwards[1]. The study, published on 3 March in *Science Advances*, is the first step in a larger project aimed at determining how the dogs have adapted to survive in one of the most radioactive places on Earth. Researchers hope to use the knowledge gained to better understand the effects of long-term radiation exposure on human genetics and health.

"We have so much to learn from these animals," says Elaine Ostrander, a geneticist at the US National Institutes of Health in Bethesda, Maryland, and co-author of the study. "This is a golden opportunity to see what happens when generations of large mammals live in a hostile environment."

## The aftermath

The immediate impacts of the accident at Chernobyl were obvious: around 30 people who worked at the power plant and firefighters who attended after the disaster died of radiation poisoning within a few months of the catastrophe, according to the World Health Organization. And in the surrounding areas, pine trees withered and many insect species vanished, unable to survive in the radioactive soil.

What is less clear is how low levels of lingering radioactive material from the disaster affect the plants and animals around Chernobyl today. A handful of studies have reported unusually high genetic mutation rates in barn swallows[2] and fruit flies[3] in the vicinity of the reactor, which is now entombed in a steel and concrete sarcophagus.



Hundreds of feral dogs live at the Chernobyl power plant and in the surrounding area.Credit: Clean Futures Fund+

However, the health effects of low-levels of radiation are still hotly debated. This matters because people risk exposure to low doses of radiation in all sorts of contexts, including through certain medical scans or while working at nuclear power plants, says David Brenner, a radiation biophysicist at Columbia University in New York City who was not involved in the research. “It’s really difficult to figure out the effects” of this type of exposure, he adds, “but pretty important that we do so”.

This was a motivating factor for co-author Timothy Mousseau, an evolutionary ecologist at the University of South Carolina in Columbia. In 2017, Mousseau joined a volunteer mission to provide veterinary care to the hundreds of stray dogs living in the exclusion zone, a 2,600-square-kilometre area around the power plant to which Ukrainian officials restrict access for safety reasons.

Over the course of three years of trips to the area, Mousseau and his colleagues collected blood samples from around 300 dogs living at the power plant and around the mostly deserted city of Chernobyl after volunteers had sedated the animals with tranquilizer darts.

DNA analysis of the canines revealed that they were not newcomers to the area. By comparing the dogs' genetic profiles to those of other free-roaming dogs in Eastern Europe, the team found that the canines in the vicinity of the power plant — some of which are related to shepherd breeds — have been isolated from other dog populations for decades. And the researchers learnt that, despite Soviet concerns during the 1980s that the dogs would migrate and spread radioactive material, most of these animals hadn't moved far: those living closest to the power plant are genetically distinct from their kin living just a few kilometres away.

# A radioactive legacy

The dogs' continued presence in the area shows that they have been able to survive and breed, even while living near the reactor, "which is remarkable", says Ostrander. The 1986 accident deposited the deadly radioactive isotope caesium-137 at levels 10–400 times higher near the power plant than in the city of Chernobyl, just 15 kilometres away.



[Massive study of pet dogs shows breed does not predict behaviour]

The canine DNA samples "are incredibly valuable" because dogs tend to share many of the same spaces and diets as humans, Ostrander says. "We've never had an opportunity to do this work in an animal that reflects us as well as dogs."

But teasing out which genetic changes in the dogs are caused by radiation and which are caused by other factors — such as inbreeding or non-radioactive pollutants — won't be easy, Brenner cautions. The team acknowledges these challenges, but the researchers argue that their detailed knowledge of these dogs' ancestry, as well as knowledge of the levels of radiation different dogs were historically exposed to, "provides an ideal focus group for our future studies".

In the meantime, Mousseau is planning another sampling trip in June. The ongoing war in Ukraine hasn't stopped the group's research. But with fewer tourists visiting and leaving food scraps, Chernobyl's dogs are struggling to get by. So the team is working with a non-governmental organization to provide food to the strays, safeguarding the survival of Chernobyl's dogs — and their radioactive legacy — in the lean times ahead.

*Nature* **615**, 384-385 (2023)

*doi: https://doi.org/10.1038/d41586-023-00629-6*

# References

1. Spatola, G. J. *et al. Sci. Adv.* **9**, eade2537 (2023).
2. Ellegren, H., Lindgren, G., Primmer, C. & Møller, A. P. *Nature* **389**, 593–596 (1997).
3. Hancock, S. *et al. Environ. Res.* **172**, 333–337 (2019).

---

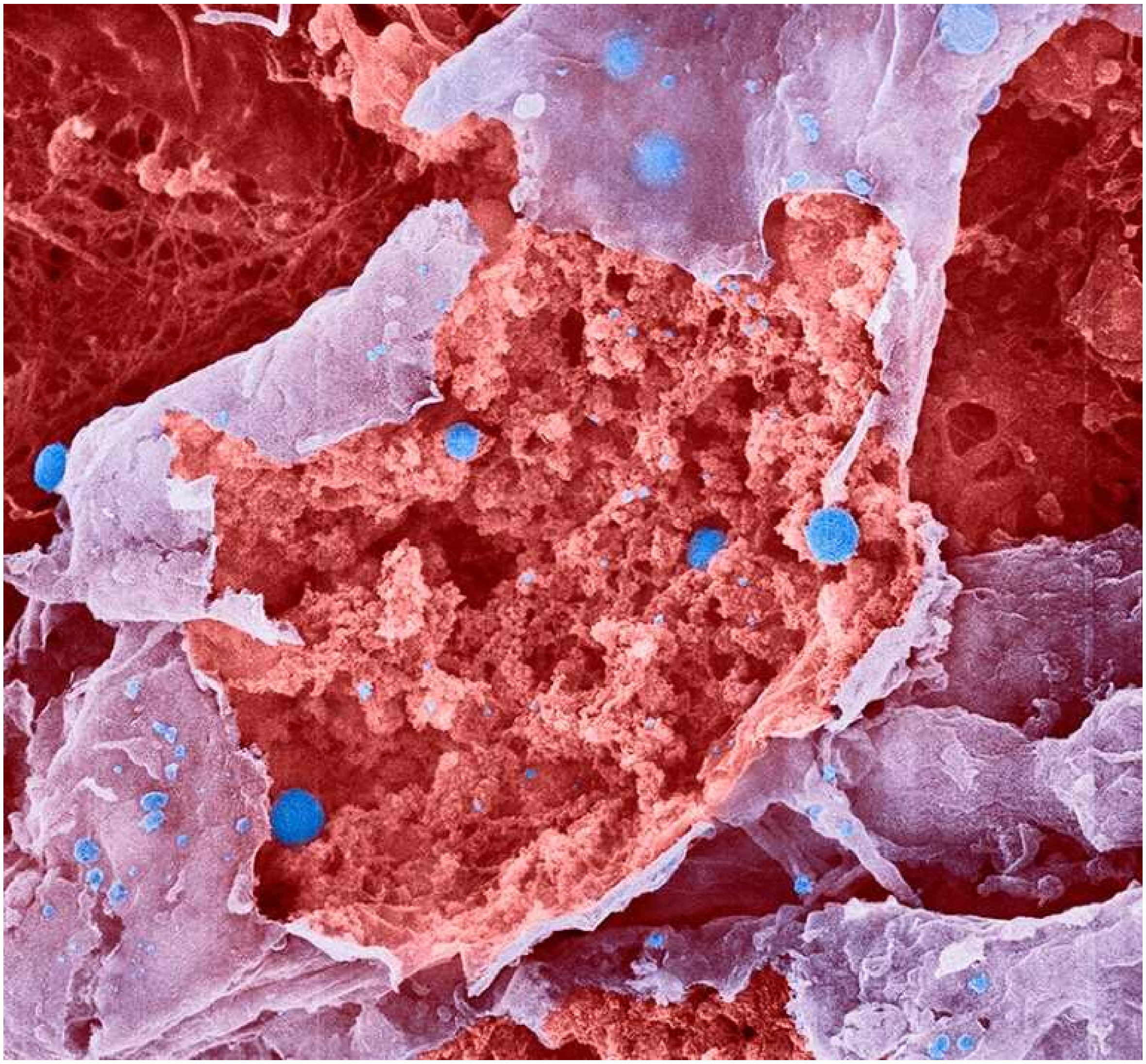
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00629-6

- NEWS
- 08 March 2023

# How the brain senses a flu infection — and orders the body to rest

Scientists trace the neurons in the throat that detect signs of infection and relay this information to the brain.

- Liam Drew



Particles of influenza virus (blue; artificially coloured) bud from a burst cell.Credit: Lennart Nilsson, Boehringer Ingelheim International GmbH, TT/SPL

A case of influenza can make even the toughest people take to their beds and lose their appetites. Now, scientists have identified neurons in mice that notify the brain of a flu infection, triggering decreases in movement, hunger and thirst[1].

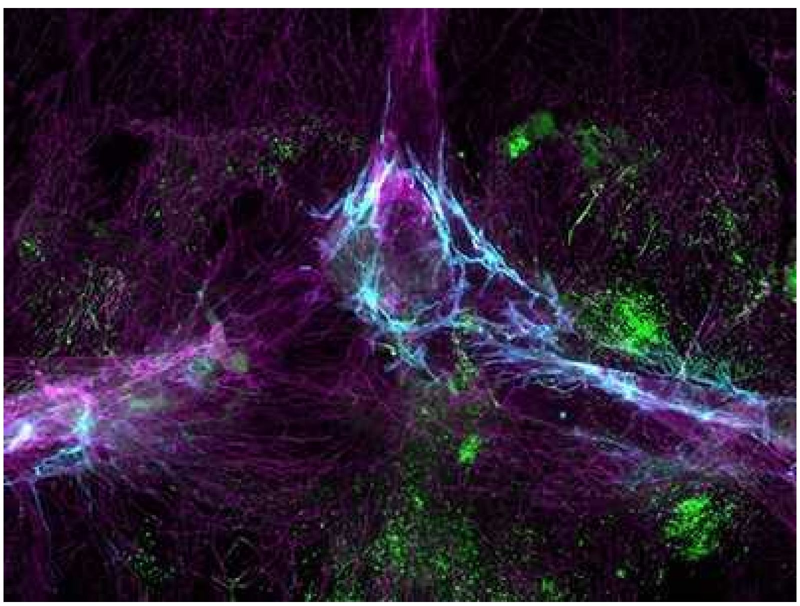
Similar neurons connecting to other parts of the body might notify the brain of other infections, too, the authors say. The work was published on 8 March in *Nature*.

“This study flips previous thinking on its head,” says Ishmail Abdus-Saboor, a sensory biologist at Columbia University in New York City who was not involved in the research. “This is paradigm-shifting in terms of how we think about sickness behaviour.”

## Brain surveillance

Before this research, “it was not clear how the brain becomes aware that there’s an infection in the body”, says study co-author Stephen Liberles, a neuroscientist at Harvard Medical School in Boston, Massachusetts. Scientists generally thought that messenger molecules from the site of infection move through the bloodstream to the brain, diffusing into it to directly activate the regions that kickstart the sickness-behaviour program.

Among the top candidates for these messenger molecules were signalling chemicals called prostaglandins, which are made in infected tissues. Aspirin and ibuprofen block prostaglandin production — and also suppress sickness behaviours, hinting that prostaglandins are key to triggering such behaviours.



Guardians of the brain: how a special immune system protects our grey matter

The authors showed that a specific prostaglandin receptor, called EP3, is responsible for generating sickness behaviours. EP3 is found on neurons throughout the body, including the brain. To test its function, the researchers deleted the brain's EP3 receptors in mice and infected the animals with flu virus. The mice still changed their behaviour — indicating that the brain is not getting infection dispatches from blood-borne prostaglandins.

Instead, the authors found that the key agents are a specific EP3-containing population of neurons located in the mouse's neck. These neurons have branches that stretch from the mouse equivalent of the tonsils to the brainstem. This geography makes sense: the tonsil area "serves as the interface between the outside air and what goes in the airway", says study co-author Na-Ryum Bin, a neurobiologist also at Harvard. The area is rich in immune cells that churn out prostaglandins when they encounter pathogens.

The results tell a narrative of illness: flu viruses enter the airway and infect throat cells, triggering prostaglandin production, and these previously unappreciated neurons respond. The infection alert then travels along the neurons' branches on "a dedicated highway to the brain", Abdus-Saboor says.

## A highway to the brain

Neural pathways do something blood-borne signals cannot: they give the brain information about exactly where the infection is occurring. The authors note that many other types of neuron have receptors for prostaglandins and other immune-related signals. They suggest that further dedicated pathways could exist, including ones for detecting gut infections, triggering nausea.

[Flu and colds are back with a vengeance — why now?]

The study also revealed a paradox. Scientists assume that there is an evolutionary advantage to sickness behaviour. But when the team blocked those behaviours, such as food avoidance, mice were less likely to die of the flu. Liberles speculates that this behaviour-modifying system might have evolved because it is beneficial in most cases of infection — even if it isn't

in all. Alternatively, behaviours such as immobility might be advantageous by reducing the spread of pathogens through populations.

The new results don’t tell the full story. The infection-sensing tonsil neurons are responsible for sickness behaviour only during a flu infection’s first stage, which affects the upper airway and lasts roughly a week. As the virus moves into the lower respiratory tract over the course of the illness, another nerve pathway takes over the job of driving sickness behaviours. “If we could find a way to block that second pathway, that, in combination, could have tremendous clinical impact,” Liberles says.

*Nature* **615**, 385 (2023)

*doi: https://doi.org/10.1038/d41586-023-00675-0*

## References

1. Bin, N.-R. *et al.* *Nature* https://doi.org/10.1038/s41586-023-05796-0 (2023).

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00675-0

- NEWS Q&A
- 07 March 2023

# Is nuclear war more likely after Russia’s suspension of the New START treaty?

The nation has spurned a major arms-control agreement, pushing nuclear powers towards a worrying lack of regulation, says non-proliferation researcher.

- Jeff Tollefson



A Russian nuclear missile is displayed in Moscow’s Red Square during a military parade.Credit: Mikhail Svetlov/Getty

The world has lurched a step closer to the prospect of nuclear war, say researchers, after Russia declared last month that it would suspend its participation in its last major nuclear-arms treaty with the United States.

Russian President Vladimir Putin's announcement will formally end inspections under an agreement known as the New START treaty (Measures for the Further Reduction and Limitation of Strategic Offensive Arms). The inspection programme had allowed scientists on both sides to verify that the other was complying with the agreement. The move comes amid soaring tensions between Russia and the West over the war in Ukraine.

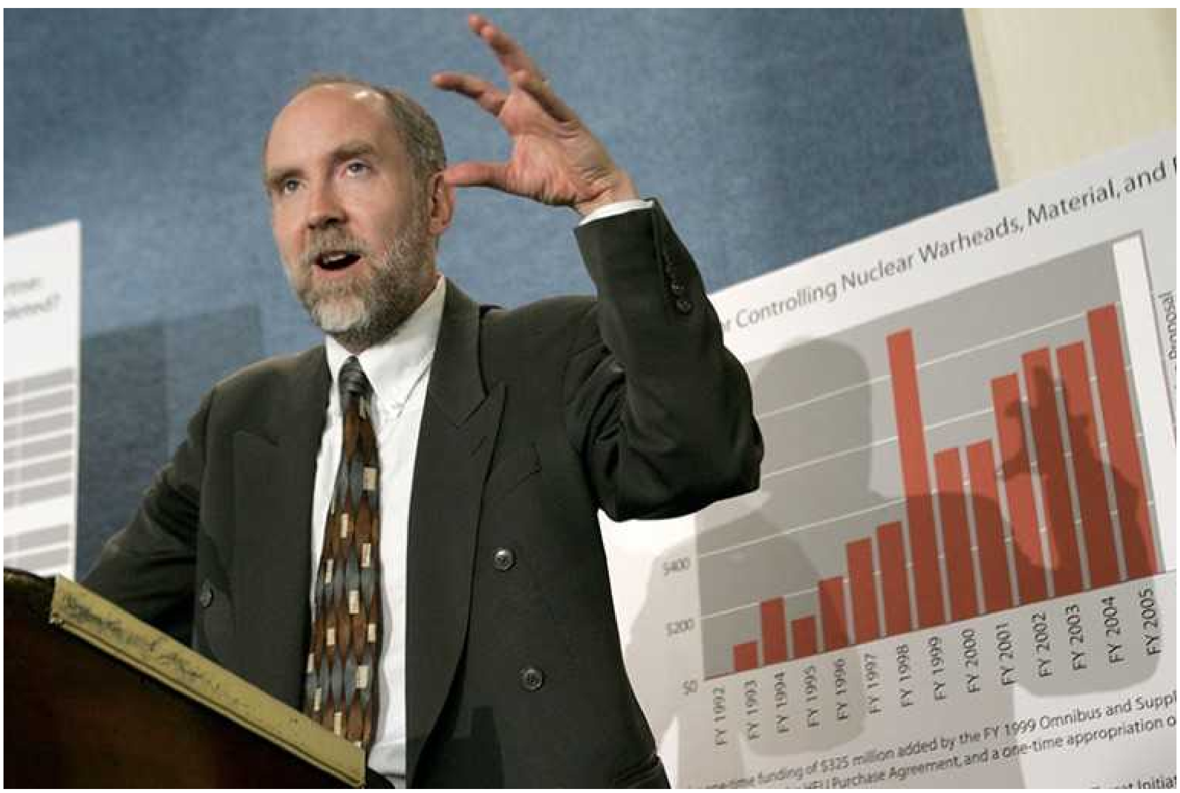


How a small nuclear war would transform the entire planet

New START limits Russia and the United States to 1,550 strategic warheads each, but has been facing challenges for years. The COVID-19 pandemic forced a pause in weapons inspections, and they have not restarted since Russia invaded Ukraine last year. US government officials said in January that Russia was violating the treaty because it would not allow inspections to resume. However, Russia's official suspension of the inspections is another blow to international efforts to control nuclear weapons, say experts, who worry that the world's largest nuclear powers will not be subject to any arms-control obligations once New START expires in three years — or even sooner.

*Nature* discussed the rising nuclear threat with Matthew Bunn, who tracks nuclear policy at Harvard University in Cambridge, Massachusetts. Bunn participates in one of the last remaining back-channel dialogues between Russian and US scientists, a mechanism that researchers developed during the cold war to foster communication — and avoid conflict.

He has spent decades cultivating relationships with Russian scholars, nuclear scientists and military officials, only to see many of those relationships evaporate since the war in Ukraine, which Bunn says has created a level of hostility between the two nuclear superpowers that has not existed since the Cuban Missile Crisis in 1962. The situation is becoming increasingly precarious across the board, he says, but there are still reasons for hope.




Nuclear-policy researcher Matthew Bunn has spent decades cultivating relationships with Russian scholars, nuclear scientists and military officials.Credit: Chip Somodevilla/Getty

## What is the danger of Putin suspending Russia's participation in New START?

It’s another important step undermining the future of arms control. The New START treaty expires in February 2026, and with the war in Ukraine, no talks are under way on any kind of replacement. Now even the New START treaty may fall apart before February 2026, and we may find ourselves for the first time in half a century in a world with no limits on US and Russian nuclear forces. That would increase dangers to all of us.

## Back-channel dialogues during the cold war helped to defuse tensions. Are they still happening?

Almost all US–Russian communication is cut off, and many non-government dialogues have been cut off. Russia appears also to have instructed current and retired people involved in its nuclear-weapons complex not to participate in these dialogues, so the kinds of conversation that I used to have regularly with Russian colleagues are just not happening anymore.



Seven ways the war in Ukraine is changing global science

However, the national academies of Russia and the United States have continued their nuclear arms-control dialogue, in which I am a participant.

We've explored a lot of the key issues that would be required to be addressed in the next round of nuclear arms-control negotiations. I won't say those discussions have been easy, and we have not found it very productive to spend much time talking about who's right and who's wrong in the war, but that dialogue is ongoing and both governments have asked us to continue.

## How should the United States and other nuclear powers respond to Russia's decision?

The United States needs to continue to maintain the New START Treaty as much as possible. It needs to get Russia back to the verification table, and I think we should do our best to maintain at least the basic numerical limits of New START, which Russia says it's also willing to maintain. And we need to find ways to get back to the table not only on arms control, but on military-to-military dialogues and other measures to try to reduce the risk of inadvertently stumbling into a nuclear conflict that no one wants.

## Russia is just one country, and New START is just one treaty. What about the rest of the world?

If you look at the nuclear landscape, you have more hostility between the United States and China. You have North Korea with a burgeoning nuclear arsenal, you have India and Pakistan with growing nuclear arsenals and you have Iran, with the collapse of the Iran nuclear deal in 2018, right at the edge of a nuclear-weapons capability. It's a much darker nuclear picture than it was even a decade ago.



Will war in Ukraine mark a new era for European defence research?

Now, does that mean everybody should run for the hills? No. I think that core deterrence is very strong. The United States and Russia, and the United States and China, have enormous incentives not to end up lobbing nuclear weapons at each other's homelands. But there's a lot to do to repair the nuclear order internationally.

## Are you hopeful?

I'm worried, but I am one of the last standing optimists. If I had to bet, I would bet that ten years from now, we will have no new nuclear-armed states. Only 5% of the states in the world have nuclear weapons, and that was also true 35 years ago. That's an amazing public-policy success story, but there is a lot of work to do and a lot of rethinking to do to reduce the nuclear dangers that have been growing in the world in recent years.

*Nature* **615**, 386 (2023)

*doi: https://doi.org/10.1038/d41586-023-00679-w*

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00679-w

- NEWS FEATURE
- 06 March 2023

# What Turkey's earthquake tells us about the science of seismic forecasting

Geologists knew decades ago that a quake would strike southeastern Turkey, but precise prediction is still the stuff of science fiction.

- Shannon Hall



The magnitude-7.8 earthquake in Turkey last month destroyed many buildings, such as this one in the city of Kahramanmaraş.Credit: Adem

Altan/AFP via Getty

Two decades ago, John McCloskey drew a red line on a map of southeastern Turkey to pinpoint where a large earthquake would probably strike. The only question was when.

The answer came last month, when a magnitude-7.8 shock hit the precise location that McCloskey and his team had identified. It struck at 4.17 a.m. local time on 6 February, when most people were asleep, and killed more than 50,000 residents in Turkey and neighbouring Syria.



One MRI for 4.7 million people: the battle to treat Syria's quake survivors

McCloskey's work shows both the promise — and limitations — of the science of earthquake forecasting. Although geologists have long attempted to provide warnings of the location, magnitude and exact time of future quakes, decades of research have shown that it's probably impossible to predict when a geological fault will start to shake. "When you try to winnow it down to know what's going to happen next, it tends to be a lesson in humility," says Susan Hough, a geophysicist in the Earthquake Hazards Program at the United States Geological Survey (USGS). "The real focus in most of the world is not on prediction, but on assessing the hazard and the long-term rates of earthquakes."

Today, researchers work on forecasting: identifying which fault segments are most dangerous and what size earthquakes they are expected to produce. Armed with that knowledge, policymakers can take steps to reduce death and destruction by, for example, requiring better building practices or urging local residents to prepare. Some regions of Japan, the United States and Turkey have developed early-warning systems that alert residents when an earthquake has started nearby. “In principle, you can get rid of seismic risk,” McCloskey says.

## Danger zone

Turkey is a seismically active junction at which several pieces of Earth’s crust meet and grind against each other. In southeast Turkey and northern Syria, the Arabian plate is pushing north against the Anatolian plate, squeezing it to the west. But the shift isn’t one smooth movement. Instead, friction holds the plates in place, sometimes for centuries. When the stress overcomes the friction, the plates on either side of the fault line will shudder past each other, releasing tremendous energy in the form of an earthquake.

This has happened time and time again in Turkey — a history that allowed McCloskey and his colleagues to map the stresses along one of its major quake sources, the East Anatolian fault. Like other faults, it is divided into segments that slip at different times. When one segment shifts and shakes, it alters the stress on neighbouring sections of the same fault and other faults nearby. That increases the stress in some places, bringing them closer to failure, but relaxes stress on others — making them safer for the time being.



A fracture cuts across a road in the Kahramanmaraş region of Turkey after two strong earthquakes on 6 February.Credit: Utku Ucrak/Anadolu Agency via Getty

“They are not just randomly occurring earthquakes,” says Ross Stein, chief executive of Temblor, a company specializing in seismic hazard and risk assessment. “They are in a conversation. And that conversation is carried out through stress transfer.”

In 2002, McCloskey (now a geophysicist at the University of Edinburgh, UK) and his colleagues used this technique to diagnose regions on the East Anatolian fault that were highly stressed. With the help of historical records, the team incorporated the stress changes caused by ten earthquakes since 1822 into a model of ongoing plate movement. The modelling suggested that a region of the fault line south of Kahramanmaraş — the precise location and length of the fault that ruptured on 6 February — was at a heightened risk of giving way at some point in the future[1]. The team even knew that it would be devastating, forecasting a quake of magnitude 7.3 or higher. “The correspondence is remarkable,” McCloskey says.

It isn't the first time that this method, technically known as Coulomb stress transfer, has accurately pinpointed an upcoming trembler. In 1997, Stein and his colleagues analysed the earthquakes that had already struck Turkey's North Anatolian fault to estimate that the next might occur near the city of Izmit[2]. Two years later, that quake arrived — killing more than 17,000 people. In 2005, McCloskey and his colleagues calculated that the shift in stress after the 2004 Sumatra–Andaman quake in Indonesia might cause one in the Sunda trench west of Sumatra[3]. It came 12 days after the study was published. And in 2008, Shinji Toda from the Geological Survey of Japan in Tsukuba and his colleagues projected that the Wenchuan earthquake earlier that year in China would increase the stress of three adjacent faults[4]. In the following decade, two of those faults unleashed powerful earthquakes.

## Added stress

It isn't possible to use the technique everywhere. Because this model requires some knowledge of previous earthquakes, often centuries in the past, researchers can use it to assess only regions where the seismic history is well known. So it is most successful in forecasting aftershocks, which are typically smaller than main shocks. Still, there are many unknowns, and scientists are working hard to evaluate the model further.

In 2002, Tom Parsons, a geophysicist at the USGS, analysed more than 2,000 earthquakes with magnitudes of greater than 5.5 that occurred after — and near — quakes larger than magnitude 7. He found that 61% of the later quakes were associated with an increase in stress caused by the earlier ones[5]. The findings suggest that Coulomb stress transfer can accurately identify faults that are more likely to cause damaging quakes, he says. Then, in 2008, Parsons and his colleagues published a forecast following the Wenchuan earthquake with the intention of later evaluating the model's performance[6]. That work is ongoing.



Why the Tongan eruption will go down in the history of volcanology

Today, Stein, one of the researchers who developed the theory about how forces shift after earthquakes[7], estimates that the method has been used in 30,000 papers to explain two-thirds of our planet's recent aftershocks and progressive main shocks. "That tells us this is not the only game in town," Stein says. "Faults are grungy, messy features and they don't behave as we would like them to."

McCloskey's model, for example, anticipated the location of the recent Turkey earthquake, but the shaking started on a much smaller branch of the fault and then spilt over to the main part, a pattern that Stein finds baffling. Another complication is that the main earthquake was also much larger than anticipated, probably because it re-ruptured a segment to the south that broke in 1822 and a segment to the north, which broke in 1893.

"This really underlines the problem of earthquake forecasting," McCloskey says. "Even when we identify the place that is most dangerous, every earthquake is unique."

Not long ago, seismologists thought they might be able to predict some quakes days or hours before they strike. Such hopes emerged from Parkfield, California, where earthquakes had rocked a small part of the San Andreas fault nearly every 22 years. Each of these quakes followed a smaller shock

to the north. And hours before a strong quake near Parkfield in 1966, precursory movement had broken an irrigation pipeline that crossed the fault.

“In 1966, earthquake prediction looked like it was ours to have,” Stein says. Before the next anticipated earthquake, geologists wired the area with hundreds of seismometers — hoping to find some harbinger that could be used to forecast future quakes. But when the next quake hit, researchers saw no warning signs.

Other precursors have similarly vanished. Over the years, scientists have analysed increasing amounts of radon in local water, electromagnetic signals from Earth’s crust and even odd animal behaviour. But none of these potential precursors stood up to statistical tests. “Despite all kinds of startling, promising shreds of evidence, we haven’t made an iota of progress toward actually predicting earthquakes,” Stein says.

McCloskey doesn’t think that it will ever happen. And Hough, who wrote a book called *Predicting the Unpredictable* (2009), argues that most geologists in the West don’t even work on it — at least, not any more. “We know how unlikely it is that suddenly something is going to show up that we can see before every big earthquake,” Stein says.

Even though geoscientists can’t predict quakes with any precision, many researchers say it is possible to prevent much of the death and destruction from these natural disasters.

After the 1999 earthquake in Izmit, Aykut Barka, a geologist at Istanbul Technical University, warned that the increased stress could trigger a similar rupture near Düzce, a town roughly 100 kilometres east[8]. His work persuaded the authorities to close school buildings that had been damaged by the Izmit shock. When a magnitude-7.1 earthquake struck the city 2 months later, the buildings collapsed.

## Early warnings

Earthquake forecasting could help in other regions as well. California, for example, which is home to the massive San Andreas fault, has implemented the beginnings of an early-warning system that relies on networks of seismometers to detect the very start of a quake. That can provide seconds or minutes of advance notice to Californians to ‘drop, cover and hold on’ while automatically triggering life-saving measures such as slowing trains to a stop.



How dangerous is Africa’s explosive Lake Kivu?

In 2002, Turkey implemented an early-warning system in Istanbul that will slow trains, open lift doors and shut down critical processes in factories in the case of an earthquake. The country has also implemented building codes, but many scientists were concerned that they weren’t being enforced rigorously enough. Mustafa Erdik, a retired civil engineer at Boğaziçi University in Istanbul and president of the Turkish Earthquake Foundation, agrees that this was the case — arguing that ignorance, incompetence and implicit collusion between architects, inspectors and builders were at fault.

That makes February’s aftermath particularly painful for those researchers who have been sounding the alarm for years. “You put a red line on a map, and you understand that means lots of people are going to be killed and their houses destroyed,” McCloskey says.

“The Turkey earthquake to me is, of course, a complete tragedy,” he says. Yet McCloskey is hopeful that we will learn from it. If we do, the next red line he draws on a map will not necessarily equal a catastrophic loss of lives.

*Nature* **615**, 388-389 (2023)

*doi: https://doi.org/10.1038/d41586-023-00685-y*

# References

1. Nalbant, S. S., McCloskey, J., Steacy, S. & Barka, A. A. *Earth Planet. Sci. Lett.* **195**, 291–298 (2002).
2. Stein, R. S., Barka, A. A. & Dieterich, J. H. *Geophys. J. Int.* **128**, 594–604 (1997).
3. McCloskey, J., Nalbant, S. & Steacy, S. *Nature* **434**, 291 (2005).
4. Toda, S., Lin, J., Meghraoui, M. & Stein, R. S. *Geophys. Res. Lett.* **35**, L17305 (2008).
5. Parsons, T. *Geophys. Res. Lett.* **107**, 2199 (2002).
6. Parsons, T., Ji, C. & Kirby, E. *Nature* **454**, 509–510 (2008).
7. Stein, R. *Nature* **402**, 605–609 (1999).
8. Barka, A. *Science* **285**, 1858–1859 (1999).

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00685-y

# Opinion

- **Five steps to make MRI scanners more affordable to the world** [14 Mar 2023]
  • Fifty years since the basis of magnetic resonance imaging was published, MRI scanners remain expensive — and impractical in many countries. Here's how we are making them smaller and less costly.

- COMMENT
- 14 March 2023

# Five steps to make MRI scanners more affordable to the world

Fifty years since the basis of magnetic resonance imaging was published, MRI scanners remain expensive — and impractical in many countries. Here's how we are making them smaller and less costly.


- Andrew Webb &
- Johnes Obungoloch




Johnes Obungoloch (centre) and his team with a magnet used in a low-field MRI scanner. Credit: Wouter Teeuwisse and Tom O'Reilly

In 2014, paediatric neurosurgeon Steven Schiff approached our research team with a challenge: build an easy-to-use, portable magnetic resonance imaging (MRI) system at a fraction of the cost of a conventional scanner. He needed it to image the brains of children in Uganda with hydrocephalus, a build-up of fluid in the brain.

Hydrocephalus is the most common reason for neurosurgery in children worldwide; there are almost 200,000 cases each year in sub-Saharan Africa alone. Surgeons operate to divert the fluid but procedures often fail and follow-up imaging is needed. Other forms of scanning and monitoring are impractical for this condition: ultrasound cannot penetrate the skulls of children over the age of two and X-rays carry a radiation risk. MRI fits the bill.

MRI is an indispensable part of clinical care, with more than 100 million scans performed worldwide annually, and around 50,000 machines in hospitals and clinics. However, people in low- and middle-income countries (LMICs) have limited access to this technology, despite constituting more than 70% of the world's population. For example, on average there are 0.7 MRI scanners per million people in Africa, compared with 55 per million people in Japan, 40 in the United States and 35 in Germany. Most systems are in big cities, distant from rural populations. Of the scanners in Africa, 39% use obsolete hardware and software[1].



One MRI for 4.7 million people: the battle to treat Syria's quake survivors

MRI scanners are expensive — it costs several million US dollars to purchase, install, maintain and operate one machine. They require a superconducting magnet a couple of metres long, which weighs 5–10 tonnes, cooled by liquid helium and housed in a shielded room to reduce electromagnetic interference. The scanners need a powerful energy source and chilled water to cool other components. Highly trained technicians are essential. Software and hardware upgrades cost tens to hundreds of thousands of US dollars each year. All of this is beyond the reach of research, education and health-care settings in LMICs.

What these nations need instead, are completely redesigned MRI systems that can be easily set up and run in rural settings and secondary health facilities. Local physicians need to have confidence in using them to diagnose and treat common life-threatening conditions. Since our first conversation with Schiff (who is based at Yale School of Medicine in New Haven, Connecticut), our groups at the Leiden University Medical Center in the Netherlands and Mbarara University of Science and Technology in Uganda have collaborated with researchers in Paraguay, Spain, the United States and Germany to develop and promote affordable MRI systems in LMICs.

Here we outline progress and highlight five key challenges on which researchers should concentrate. We focus on Africa, the region with which we have most experience, but similar situations exist in South America and South Asia.

## Improve low-field technologies

The principle behind MRI was published 50 years ago this week in *Nature* by Paul Lauterbur[2], a chemist from the United States who shared the Nobel Prize in Physiology or Medicine in 2003 with Peter Mansfield, a British physicist, for this work. Conventional MRI machines use powerful superconducting magnets to polarize hydrogen nuclei in tissue. Pulses of radio waves interact with the polarized nuclei, generating a voltage in detectors placed around the patient. Algorithms convert these signals into images in which physicians can distinguish between tissues on the basis of characteristics such as water or protein content and stiffness.

A compact, portable MRI system cannot be built by simply shrinking down existing designs or replacing components with cheaper alternatives. But small machines can be optimized to monitor specific conditions such as hydrocephalus.



Scientists at the Mbarara University of Science and Technology in Uganda.Credit: Wouter Teeuwisse and Tom O'Reilly

To start, the superconducting magnet can be replaced with a small permanent magnet, which is typically made of neodymium boron iron or samarium cobalt alloys. Such magnets produce a much weaker field (0.05 to 0.1 Tesla) than conventional machines (1.5 or 3 Tesla), and hundreds to thousands of times less signal.

Electrical components must be redesigned to run on less power and off mains electricity or batteries. Image-processing algorithms and hardware must be able to correct for external noise, so that no shielding is needed. And the software must be easy to operate with minimal training.

Over the past ten years, two types of low-field MRI system have been developed by various groups. We have designed a version that consists of thousands of tiny magnets (a few centimetres or less across) configured to optimize the strength and uniformity of the magnetic field around the

patient. The system, a barrel roughly 50 cm long and 50 cm in diameter, with an open inner bore 31 cm across, is relatively light (around 75 kilograms) and can be assembled in modules without specialist tools or equipment[3]. The system can be positioned at the head of a hospital bed and the patient slid into it. Our scanner can be wheeled through a hospital, loaded into a car and operated using open-source software.

The second design is simpler and consists of two discs that are a few centimetres thick and 60–95 cm in diameter. The discs are made of magnetic material joined by one or two iron yokes, placed above and below the patient's head[4]. This system, however, requires specialized equipment to assemble, weighs several hundred kilograms and has to be moved on a motorized carriage.



Could Africa be the future for genomics research?

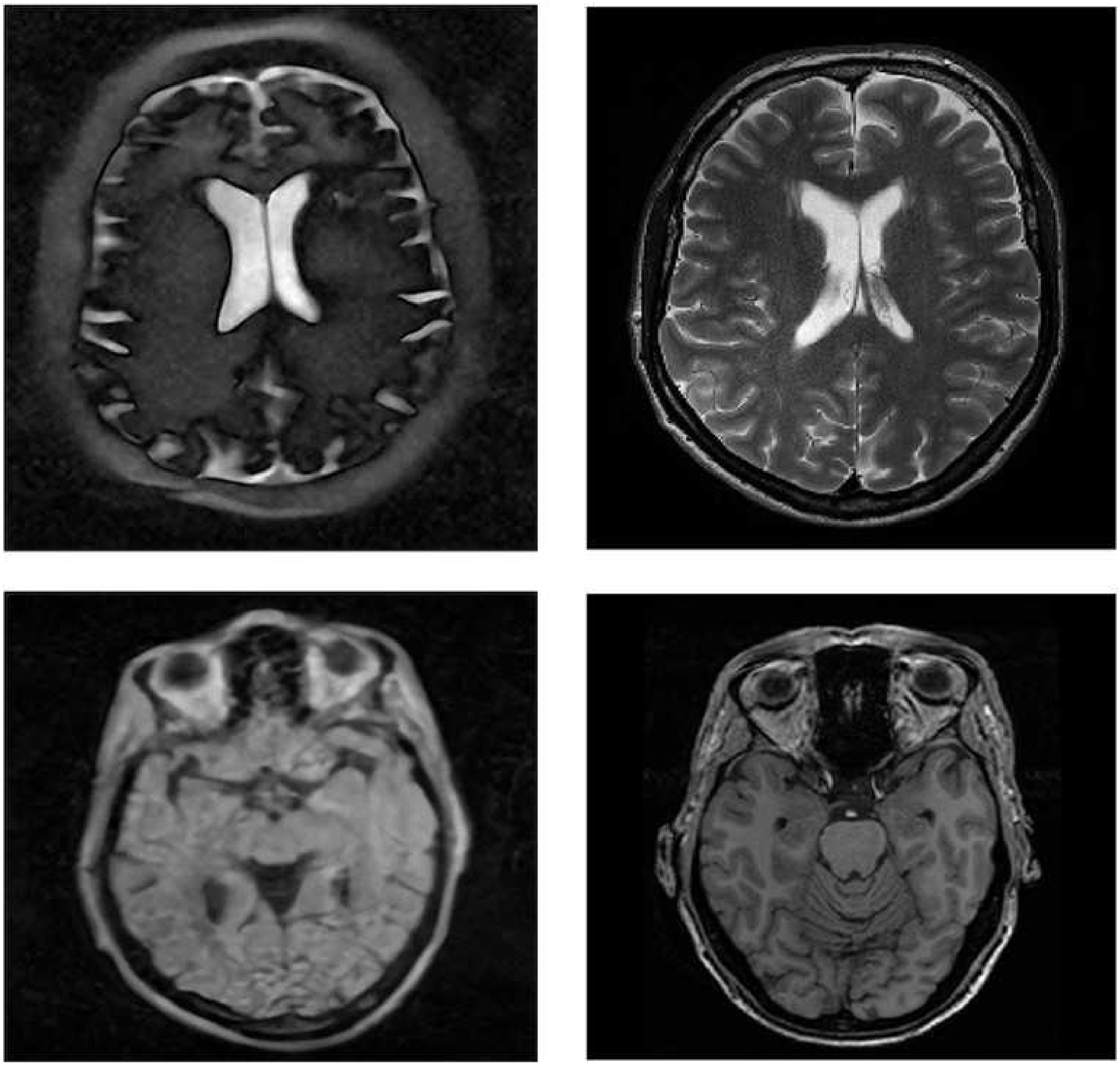
The US company Hyperfine in Guilford, Connecticut, has commercialized this latter design, which is currently used alongside typical MRI systems in intensive-care units and emergency departments, mainly in the United States. It has been tested on people with brain haemorrhages, strokes, multiple sclerosis and brain injuries[5]. The Bill & Melinda Gates Foundation plans to place more than 45 such systems throughout LMICs to help researchers understand and address the impacts of maternal and infant health, malnutrition and disease on infant brain development. Similar research platforms have been developed in China and Hong Kong.

Currently both systems are capable of imaging the brain with a resolution of around 2 millimetres in 10 minutes. By comparison, a conventional MRI scanner might resolve structures smaller than one millimetre in a few minutes. Although the low-field systems' performance is sufficient for straightforward conditions such as hydrocephalus, it can be improved. Researchers must design more sensitive detectors, increase image contrast and deploy advanced image reconstruction methods, for example, using artificial intelligence (AI). They also need to ensure that systems operate reliably, robustly and reproducibly in challenging environmental conditions that are likely to be encountered, for example, in rural settings in LMICs.

## Build physicians' confidence

Conventional MRI is a specialist practice, usually handled by radiologists who are familiar with all the nuances. For uptake in LMICs, it's crucial that physicians see low-field MRI technology as 'smart-tech' rather than 'low-tech', and not just a cheap alternative to the systems used in high-income countries (HICs).

Building this confidence requires clinicians and scientists to share their understanding of the information contained in the 'grainier' images obtained with low-field MRI. There's currently little knowledge of, and few data on, how to use these devices clinically. MRI scientists need to ensure that the quality of low-field images is robust and reproducible and that, as with conventional scanners, any non-clinical artefacts are easy to spot. Simple user interfaces need to be designed and improved by working with clinicians and technicians in LMICs.



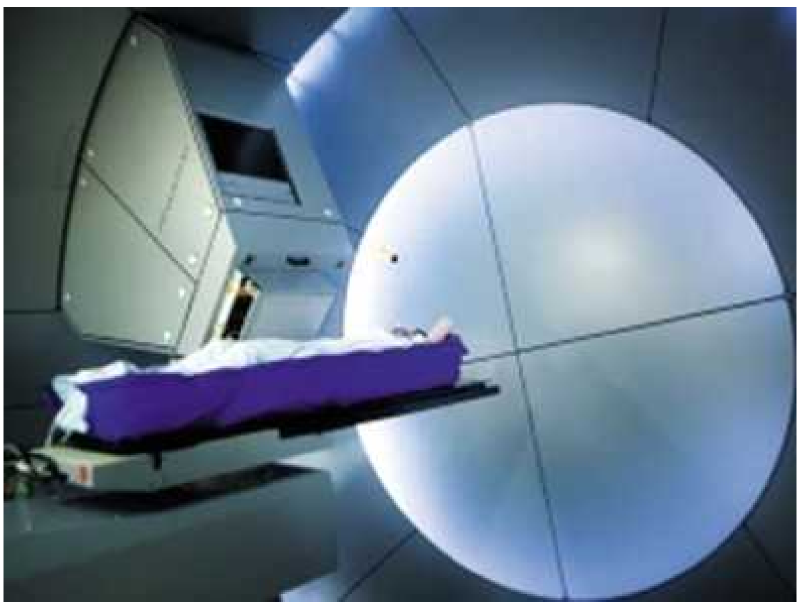
Scans from a low-field MRI system (left) compared with typical MRI scans.Credit: Leiden University Medical Center

Training programmes need to be developed for clinicians, including showing them how low-field images compare with conventional high-field and other types of scan. Acquisition of these images should be performed in both HICs and LMICs — in the former because of greater hardware availability and in the latter to ensure relevance to local health issues. The MRI community must make the case and prove that low-field images are clinically useful.

## Prioritize open-source approaches

Commercial MRI scanners use proprietary software and hardware. This limits the extent to which medical physicists and engineers can understand and repair these systems. Many places do not have such expertise available locally. The World Health Organization (WHO) estimates that 70% of medical equipment donated to LMICs cannot be used because of insufficient training.

Scientists, clinicians and medical technologists in LMICs need to be empowered to design, maintain and build low-cost MRI systems on site[6]. And to do so, they need access to open-source hardware and software, in line with the United Nations' 2020 Roadmap for Digital Cooperation, which stresses the need for digital inclusion for all, and for building architectures for digital cooperation.



Three ways to make proton therapy affordable

As an example, last year we designed a low-field MRI scanner 'kit' based on open-source hardware and software and shipped it from Leiden to our engineering college in Mbarara. Biomedical engineering students from six sub-Saharan African countries used the kit to construct the first portable, custom-built MRI system on-site in Africa, which is now being used to train students.

This project highlights the importance of funding, whether it's from agency, governmental or corporate organizations, for encouraging open-source technology. Licences that allow technologies to be distributed widely while maintaining rights for developers are emerging, as are royalty-free sublicences that require a certain proportion of systems be sold at cost in LMICs. Such agreements are allowing some pharmaceutical companies to share COVID-19 vaccine technologies in Africa, for example.

What can researchers do? They can make data and code available in repositories, such as GitHub, and ensure that code is written in open-source languages. They can document methods and hardware developments on freely available websites (such as opensourceimaging.org). And they can help to adapt the academic culture, so that transparent and reproducible research is highly valued.

## Develop AI tools for LMIC contexts

Using AI to detect simple abnormalities in medical images can improve outcomes and save costs in nations where training is hard to access. AI might also be helpful for 'task-switching', in which nurses and medical technicians take on roles typically filled by clinicians. However, most current AI applications, both commercial and research, are targeted to circumstances in HICs.

Efficient AI algorithms for low-field scanners in LMICs must be customized for specific equipment and trained on local images. Data from HICs might not be relevant to populations in LMICs, owing to different spectra of disease, genetics and environmental differences[7]. Local data privacy laws must be considered, because these vary by country[8]. Images from other types of scan (such as computed tomography) that are more widely used in LMICs can be incorporated into training data. Eventually, low-field MRI systems themselves could be instrumental in increasing the size and diversity of AI training data sets, because these devices can be taken into patients' homes[9].

## Boost funding for sustainable MRI

Collectively, we found it hard to secure funding for our compact MRI system. Governments in the United States and Europe prioritize high-end instrumentation that directly benefits their residents. Research that simplifies instruments, particularly for use in LMICs, is seen as a lower priority. Philanthropic foundations are more amenable but have their own priorities and can be hard to access without the right contacts. In our case, we managed to use a small amount of initial funding from the Dutch Research Council (NWO) as a stepping stone to further grants from the European Union and US National Institutes of Health.



A patient is slid into a portable MRI system, which is produced by the US company Hyperfine.Credit: Hyperfine, Inc.

This funding gap might stem from a lack of awareness of medical needs in LMICs. Fewer than 1% of attendees at major MRI research and radiological meetings are from institutions in LMICs[1]. Similarly, research publications on MRI from LMICs make up less than 1% of global output, with the majority coming from collaborations with scientists in HICs[1]. A research brain drain from LMICs is well documented[10]. Yet, developing low-field

MRI into a clinical tool rivals the hardest challenges in medical imaging today.

In HICs, medical-imaging research involves networks of companies, government-funded laboratories and university research and education programmes. Personnel and ideas flow between these entities, enabling sustained long-term collaborations. Similar networks need to be set up in LMICs, with early-career researchers at the heart to take these technologies forward.

There are encouraging signs of change. In 2019, the European Society for Magnetic Resonance in Medicine and Biology formed a network — the Consortium for Advancement of MRI Education and Research in Africa (CAMERA; see cameramriafrica.org). With more than 54 academic volunteers, 114 radiology residents and 54 medical physics trainees across 18 African countries, CAMERA organizes local events and supports researchers' applications for funding. In 2022, the consortium assessed MRI use across sub-Saharan Africa, highlighting gaps in training, maintenance and research capacity[1]. Larger societies are beginning to mirror such efforts, including the International Society for Magnetic Resonance in Medicine (ISMRM) and various European and US radiology societies.

To ensure that LMIC scientists are involved, we advocate that financial obstacles to society membership and conference attendance be removed. Stop-gap measures, such as online-only access and associate memberships, are not enough, because they lack the benefits offered to full members, such as networking, building community awareness and the right to vote. Major organizations need to arrange their annual conferences in LMICs on a regular basis.

As Schiff's challenge showed us, improving global health demands more than a piecemeal approach. Creative concepts from researchers and physicians can make MRI accessible to everyone.

*Nature* **615**, 391-393 (2023)

*doi: https://doi.org/10.1038/d41586-023-00759-x*

## References


1. Anazodo, U. C. *et al. NMR Biomed.* **36**, e4846 (2022).
2. Lauterbur, P. C. *Nature* **242**, 190–191 (1973).
3. O'Reilly, T. & Webb, A. G. *Magn. Reson. Med.* **87**, 884–895 (2022).
4. Liu, Y. *et al. Nature Commun.* **12**, 7238 (2021).
5. Prabhat, A. M. *et al. Front. Neurol.* **12**, 760321 (2021).
6. World Health Organization. *Local Production and Technology Transfer to Increase Access to Medical Devices* (WHO, 2012).
7. Pereira, L., Mutesa, L., Tindana, P. & Ramsay, M. *Nature Rev. Genet.* **22**, 284–306 (2021).
8. Shen, F. X. *et al. Neuroimage* **238**, 118210 (2021).
9. Deoni, S. C. L. *et al. Brain Struct. Funct.* **228**, 493–509 (2023).
10. El Saghir, N. S. *et al. JCO Glob. Oncol.* **6**, 185–189 (2020).


---


This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00759-x

# Work

- **How to wrap up research projects gracefully** [08 Feb 2023]
  • Four researchers offer tips on ending, redefining and handing over projects.
- **How cell-free processes could speed up vaccine development** [13 Mar 2023]
  • Rather than relying on yeast and bacteria, bioengineers are looking for cheaper and easier ways to make proteins and other biomolecules.
- **Snub-nosed monkeys have taught me important life lessons** [13 Mar 2023]
  • Zoologist Zuofu Xiang's research in Asia helps governments to protect the populations and teaches him the value of cooperation.

- CAREER FEATURE
- 08 February 2023

# How to wrap up research projects gracefully

Four researchers offer tips on ending, redefining and handing over projects.

- Andy Tay

Find a new job



Credit: Getty

Some research projects arrive gracefully at a natural conclusion, such as when a trainee leaves the laboratory. Others come to a screeching halt when funding runs out. Still more dance along the edge of viability, daring lab heads to pull the plug and move on. Four researchers share best practices on deciding when projects should end, and how best to wrap them up or redesign them. They advocate using approaches that ensure that colleagues' interests, feelings and career prospects are protected in the process.

## MARÍA CLARA INGARAMO: Make use of your professional circle

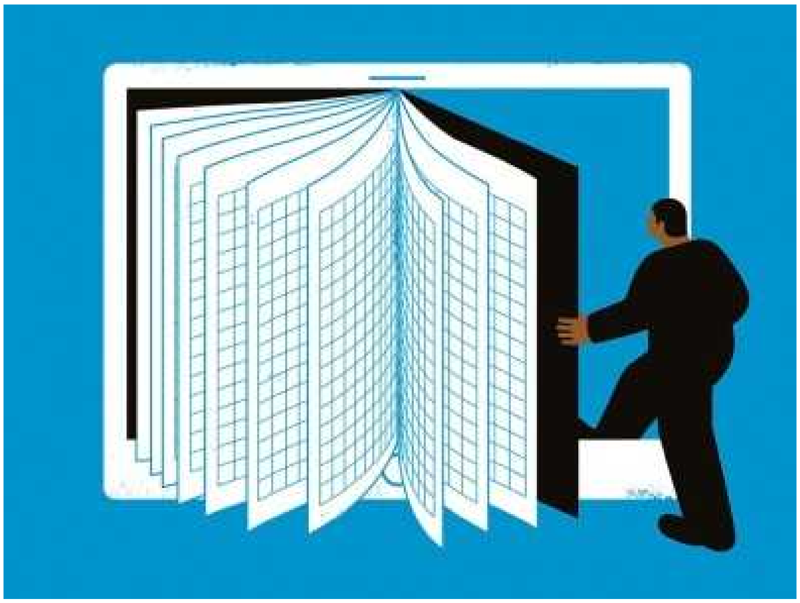
*Postdoc, National University of Litoral, Santa Fe, Argentina*.

A project can end for various reasons, including getting scooped by competitors or when funding runs out. In Argentina, we often face limited research funding compared with other regions of the world. A typical grant provides only about three million pesos (roughly US$18,800) for three years. This is often not enough to buy chemicals and pay for research staff and students. Furthermore, because Argentina does not manufacture many of the chemicals and equipment used in science, import fees can add up quickly. The weakening of the peso against the US dollar is also a problem. I know a couple of researchers from Argentina who ended their projects and headed overseas for better career opportunities.

I think ending a project can have different meanings to different people. To me, I like to think of ended projects as being deprioritized — I can always return to the idea at a later time. As a PhD student at National University of Litoral from 2016–21, I was studying the role of the *p53* gene and protein in the metabolic regulation of adipose fat tissues in *Drosophila*. However, when I collected the metabolomics data, I was overwhelmed by the plethora of regulatory pathways of enzymes. I simply did not have the right expertise to manage and interpret the huge data set.

I ended up deprioritizing the project for eight months. During this period, I consulted my fellow lab members to get support with experiments and data analysis. At the institute's weekly seminars, I asked for suggestions from colleagues studying metabolic enzymes. Eventually, I sought out a

collaborator — geneticist Norbert Perrimon at Harvard Medical School in Boston, Massachusetts — who was able to provide me with guidance. I even joined his lab group for a year to conduct more follow-up experiments. I think that, before permanently ending a project, it is useful to consult your professional circle, either to reassure you about your decision or to find alternatives. It is OK to pause projects and resume them after you acquire new knowledge, techniques and training.



How to pick an electronic laboratory notebook

In 2021, I took three months of maternity leave the last year of my PhD. I had to end my part in the project and hand it over to a colleague. To do so effectively, I shared with him my experimental protocols, organized shared folders for data and had continual e-mail, text and audio exchanges. When my maternity leave ended, I took over the same project from where he ‘ended’. Maintaining consistent communication with lab members enabled me to stay updated with the project progress and resume ownership of it when I was ready.

## XIAODONG CHEN: Redefine a project instead of ending it

*Professor of materials science and engineering, Nanyang Technological University, Singapore.*

Ending a project is not easy, because there are often multiple students and collaborators involved, each of them with different stakes in the work. If a project gets dropped, it could affect a graduate student who has spent the past three or four years working on it more profoundly than it would a senior faculty member who has multiple other projects running.

Instead of telling a student that their project has to end, I personally think it can be helpful for supervisors to explain it as 'redefining' their work. In my lab, most of our ideas might not work out, but ego or pride could deter researchers from willingly ending their projects entirely. By packaging how I assessed their projects in a more positive way, such as telling them that it is OK to publish negative outcomes, I can convince them to rethink and redefine their work to reflect what the data are suggesting, rather than clinging to their original idea. I also advise my students to talk to experts in my university and research field, and to carefully and critically examine how the project can be redefined and still remain impactful.

But when a project cannot be redefined, or must come to an end for some other reason, then I make sure to have a deep discussion with the student to explain why this is the case. In 2014, my group discovered a way to fast-charge batteries, and we even created a start-up firm to commercialize the technology. However, our start-up failed, because we were not able to manufacture the materials at the scale that companies wanted. This experience made me realize that some cannot be redefined, because the technology is not ready to make an impact.

However, this did not prevent the project from being revived. In 2016, a PhD student joined my lab and was interested in the project and in 2020, we finally found a way to scale up the production of the materials so they could be assembled into fast-charging batteries. I was able to revive the project because my student and I believed in the idea and in each other.



Molly Shoichet changed the scope of her lab's research to focus on biomedicine.Credit: Roberta Baker/University of Toronto

## MOLLY SHOICHET: Learn to let go and help transfer knowledge

*Professor of chemical engineering and applied chemistry, University of Toronto, Canada.*

Over the 27 years of my career, I have witnessed the end of many projects, often for good reasons. For example, when I started my lab at the University of Toronto in 1995, I tried looking into modifying fluoropolymers, a class of polymer that I researched for my PhD, for biomedical applications. However, I soon realized that, whereas the modified fluoropolymers were more useful for industrial applications, their biological effects were difficult

to predict or control. At the same time, the scope of my lab was evolving to focus more on health and biomedicine.

In 2009, I decided to end our fluoropolymer research completely, not because of a lack of funding, but because of my new academic interests in therapeutics delivery and drug testing. I would advise early-career researchers to explore different research areas, but be willing to end projects that are less promising or of less interest.

I've also ended projects after realizing that my approach is not practical. Glaucoma is a disease in which increased pressure in the eye results in optic-nerve damage and ultimately, blindness. I tried to engineer elastic peptides that could be injected around the optic nerve to strengthen it. Although the engineering idea was cool, I realized after talking to biologists and clinicians that the complexity of the biology meant that this approach probably would not work. I would advise students to consult various stakeholders for their opinions before starting and ending projects.

Projects also end in a lab for other reasons. In 2019, Mike Cooke (a former postdoctoral fellow in my lab) and I co-founded the company AmacaThera, based in Toronto, where I am currently the chief science officer.AmacaThera's product is an injectable hydrogel for therapeutic delivery into the central nervous system. While my research group continues to advance the fundamental biomedical-engineering aspects, the start-up is scaling up manufacturing and demonstrating clinical safety. The research 'ended' in our lab only to be handed to a commercial partner to further develop the technology for translational impact.

Knowledge retention and transfer is key to any research lab, but can be difficult when a project's lead researcher leaves. To enhance knowledge transfer, I typically try to pair an incoming first-year PhD student or postdoc with a senior lab member who will probably be leaving, such that the researchers overlap for about one year. Ideally, the pair would work closely so that, when a paper requires revision, the new lab member can take care of it. I also have technicians who provide institutional memory for previous protocols and strategies. Lab notebooks, physical or electronic, are useful resources to restart projects that ended when lab members leave and are busy with their new jobs.



[Share methods through visual and digital protocols]

## DRAGANA SAVIC: Avoid emotional attachment to ideas

*Co-founder, INIA Biosciences, Oxford, UK.*

During both my PhD and postdoctoral work, I had to end several projects. Sometimes terminating projects would be devastating if I had invested time, effort, resources and even emotions into them. Sometimes I was accountable to other partners and collaborators, too, and communication needed to be clear. For example, I was studying the role of carnitine, a molecule that transports fatty acids to the cellular organelles known as mitochondria, in various tissues. I was juggling between imaging carnitine *in vivo* — which I could do in the lab — and sending samples of extracted carnitine for *ex vivo* investigation by collaborators in another city.

Often, there were logistical constraints, because the samples had to be delivered in a timely manner. I felt that I needed to stay in control (rather than waiting for data from collaborators who had other projects) and do what I could in my lab to achieve my PhD goal in time. In the end, I decided that I would exclude the *ex vivo* data from my PhD thesis before initiating collaboration, to reach a common understanding on the expected outcomes

and timeline with the collaborators. Throughout this process, I learnt to be objective and more efficient in my PhD work, by terminating experiments that weren't panning out and moving on to a different one.

Many academic researchers are emotionally invested in their projects, which is great — you want to love doing what you do, but there is a fine balance between being persistent and being stubborn. A friend who was studying a particular drug for a few years encountered such a problem. He tested it in a preclinical pilot study, which produced mostly negative results, but he nevertheless took it to a larger study. This didn't show any particularly significant results, which led to a loss of resources and time.

It is also crucial for researchers to understand that their projects do not define them, and ending an underperforming project can be beneficial because it frees up resources for other activities.

In academia, poor-yielding projects can essentially continue if the principal investigators and researchers wish to retain them. However, in start-ups and industry, when the costs of product development are too high — when products do not generate more value to patients than existing, competing products for example, or when they fit too poorly to the market — the project will be terminated without question. Ending projects is a lot more straightforward.

In my own start-up, INIA Biosciences, my team and I originally planned to create a medical device that can both monitor and treat immune diseases. We wanted to monitor cytokine biomarkers in blood and provide a non-invasive therapy to suppress the immune system. However, after speaking to investors, vendors, patients and subject experts, we decided to focus on the therapeutic aspect and ended essentially 50% of the initial business idea on cytokine monitoring. Now INIA is in the fundraising phase. I think this strategy is useful because it allows us to present a focused business idea to funders and investors, and devote resources to the most promising idea to generate the best possible outcomes. If this ruthlessness can be practised in academia, it could greatly enhance research productivity and impact.

*Nature* **615**, 549-550 (2023)

*doi: https://doi.org/10.1038/d41586-023-00377-7*

These interviews have been edited for length and clarity.

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00377-7

- TECHNOLOGY FEATURE
- 13 March 2023

# How cell-free processes could speed up vaccine development

Rather than relying on yeast and bacteria, bioengineers are looking for cheaper and easier ways to make proteins and other biomolecules.


- Carrie Arnold




An artistic rendition of an iVAX vaccine being prepared in the field.Credit: Justin Muir

In January, the US Food and Drug Administration granted breakthrough status to a vaccine for bacterial pneumococcal disease.

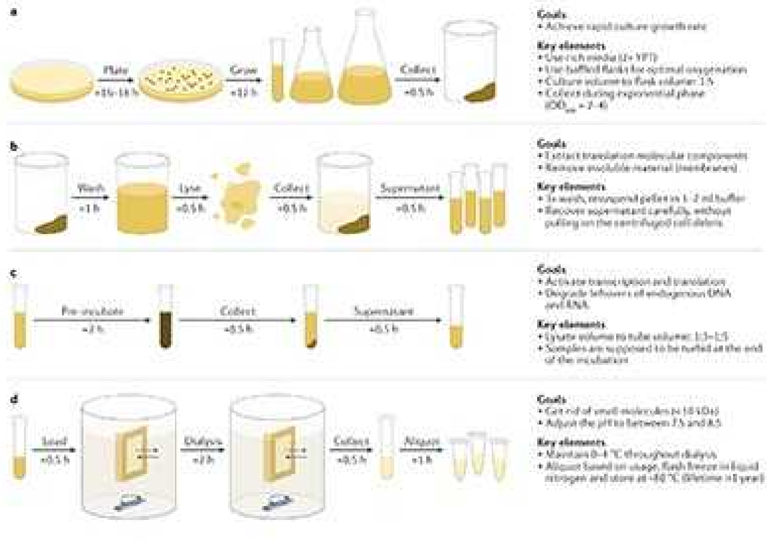
Developed by Vaxcyte in San Carlos, California, VAX-24 induces an immune reaction to *Streptococcus pneumoniae* by exposing the body to bacterial sugars linked to a carrier protein — known as a protein conjugate vaccine. Market-leading PREVNAR-20, produced by US pharmaceutical giant Pfizer, has a similar design. But whereas PREVNAR-20's protein

conjugates are purified from bacteria, those in VAX-24 are built biochemically.

Vaxcyte is one of a growing number of biomanufacturing firms embracing this cell-free biosynthesis strategy. Instead of relying on yeast or bacteria to make biomolecules, the approach strips away all the components of a cell that make it ‘alive’ — the long strings of DNA, the complex origami of the endoplasmic reticulum and even the bulwark of the cell membrane — and freeze-dries what remains. Scientists can then rehydrate the dried material with water and add nucleic acids to program the molecular machinery to make an infinite array of proteins on demand.

Tobias Erb, a microbial physiologist at the Max Planck Institute for Terrestrial Microbiology in Marburg, Germany, likens the process to using a commercial cake mix: just add water and bake. To Michael Jewett, a chemical and biological engineer at Northwestern University in Evanston, Illinois, the strategy is more like popping open the bonnet of a car, removing the engine and repurposing it to power a drill.

“It allows biology to become more chemically diverse, and at the same time, it allows chemistry to become more complex,” Erb says. Jewett has shown, for instance, that the ribosomes in cell-free systems can be used to build biopolymers with new chemical backbones, as well as proteins. Other researchers have used cell-free systems to synthesize proteins using non-standard amino acids.




[Cell-free gene expression]

Researchers have been using cell-free systems for biochemical reactions for decades — but mostly in the research laboratory. Now, thanks to advances in both reliability and scale, cell-free synthesis is emerging as a major tool for everything from diagnostic-sensor development to vaccine biomanufacturing.

Still, challenges remain. The stripped-down simplicity that makes cell-free systems so desirable also complicates key protein modifications, including the attachment of carbohydrates (glycosylation) and folding assisted by molecular chaperones. Those hurdles will need to be overcome before cell-free technologies can be used for biopharmaceutical manufacturing and other applications.

But doing so, says Matthew DeLisa, a bioengineer at Cornell University in Ithaca, New York, is really just an issue of engineering. “Cell-free systems,” he says, “turn protein synthesis into more of a chemistry problem than a biology problem.”

## Beer to biopharmaceuticals

From the natural fermentation processes that produce cheese and bread, to bioengineered cells that can pump out insulin, people have long used cells as living factories.

But efficient processes require careful balance — bioengineers must compete against the cell’s internal machinery, which is more interested in helping the cell to grow and divide than in doing industrially useful work.

“Cells don’t really want to make sustainable products. They don’t want to make insulin for us, they don’t want to act as a diagnostic. From an evolutionary perspective, that just hasn’t been their main objective,” says Jewett.



Michael Jewett works on cell-free processes at Northwestern University.Credit: Courtesy of Northwestern University

To make the process more efficient, scientists break open (or lyse) living cells to extract the molecular machinery of interest, leaving other components behind. In 1907, German chemist Eduard Buchner won the Nobel Prize in Chemistry for his work using cell-free yeast lysates to

demonstrate the biochemistry of fermentation. Around half a century later, researchers at the US National Institutes of Health used bacterial cell-free extracts to decipher the genetic code.

“You have all the benefits of working with a living system without all the problematic overhead,” says Elizabeth Strychalski, who leads the Cellular Engineering Group at the US National Institute of Standards and Technology in Gaithersburg, Maryland.

So why has biomanufacturing been slow to adopt the technology? In part, it is simply down to inertia. “It’s just taken a long time for people to wrap their heads around the idea that we don’t need a living organism to do complex metabolism, to make complex molecular rearrangements, even to make complex proteins,” says James Swartz, a chemical engineer at Stanford University in California.

## Proteins on demand

One reason for the slow transition is that it has been cheaper and easier to grow large quantities of bacteria and yeast than to run biochemical reactions, which require expensive reagents, energy sources and other materials. Cells also carry out those reactions in a defined order, something that is difficult to ensure in a vial. And they can generate higher yields of the desired product.

But now, years of molecular tinkering have made cell-free systems more efficient and cost-effective. The biggest factor behind industry’s growing interest, says Michael Nemzek, chief executive of Tierra Biosciences in San Leandro, California, is improvements in peripheral technologies such as long-DNA synthesis and miniaturization that make it easier to make longer proteins, and to test them in small batches.

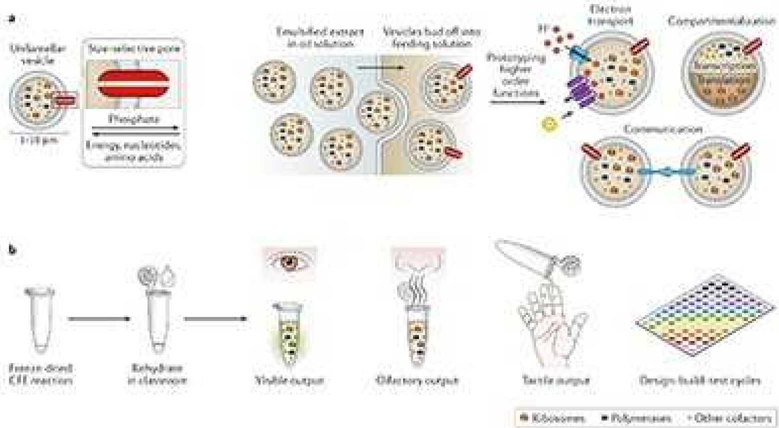
The other key advance, Jewett says, involves storage. Instead of requiring refrigeration and a complex cold chain, researchers can store most cell-free preparations indefinitely and rehydrate them as needed. “It’s kind of like freeze-dried ice cream — something you can transport wherever you go,” he says. This gives the technology added speed and flexibility — features that are all too often missing from the global vaccine supply chain.




Cell-free processes could be used to make biosensors that detect contaminants in water.Credit: Northwestern University

Despite rapid advances in messenger RNA technology, most of the world's vaccines are made of protein conjugates, which use a polysaccharide–protein duo to elicit an immune response. Theoretically, says Jewett, a lab could rehydrate a freeze-dried cell extract and begin producing such vaccines in an hour. Public-health officials wouldn't have to wait for pharmaceutical companies to ship doses from halfway around the world, says Stanford bioengineer Jessica Stark. Pharmacies and health departments could make vaccines and therapeutics on demand.

"The COVID-19 pandemic has really shone a light on the fact that we need better ways to make and distribute medicines very quickly in order to address emerging pathogens," Stark says.




[Cell-free gene expression: an expanded repertoire of applications]

Vaxcyte's partner, Sutro Biopharma in South San Francisco, California, relies on cell-free biosynthesis to design, discover and synthesize protein therapeutics for cancer, including antibodies that deliver drugs to specific cells or that recognize two antigens instead of just one, as well as immune-system proteins called cytokines. Tierra Biosciences uses cell-free biosynthesis to create everything from custom proteins for drug discovery and clinical trials to peptide libraries for screening and research. "You can upload a protein sequence, wait a few weeks, and we'll FedEx you a package," Nemzek says.

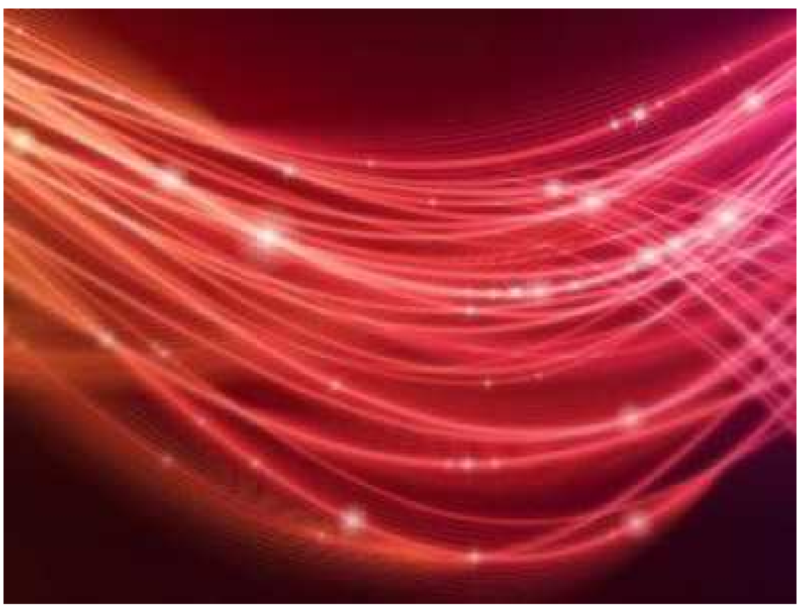
Khalid Alam, chief executive of Stemloop in Evanston, Illinois, has taken his company in a different direction. Instead of drugs, the firm uses cell-free systems to create biosensors from DNA-binding proteins called transcription factors. These proteins are switched on by molecules in the environment to induce the expression of a fluorescent or colour-coded protein that the user can see. Alam hopes for a wide range of applications, from treating infectious diseases to detecting heavy metals in drinking water — what Jewell likens to "a pregnancy test for water".

"Our mission is to unlock this power that biology has to sense and respond to the changing conditions of the environment," Alam says. The sensors "empower anybody to sense the world around them".

# Moving forward

Drawn by the possibility of building therapeutics and biosensors on demand, the US Army awarded US$13 million last March to fund a new Cell-Free Biomanufacturing Institute at Northwestern University, with Jewett at the helm. Its goal is to tackle ongoing issues of cost, yield and reproducibility to facilitate more widespread use of cell-free systems.

Antibodies, for instance, have become the bread and butter for many pharmaceutical companies, and DeLisa says that the ability to synthesize these molecules using cell-free methods would be a “game changer”. But even the simplest antibody remains too complex for existing cell-free platforms to tackle, he says. There’s no fundamental reason that the technology can’t synthesize an antibody, but scientists have yet to work out precisely how to do so.



NatureTech hub

Part of the problem is folding: many complex proteins require other proteins called chaperones to attain their final form.

There’s also the sticky issue of glycosylation. Half of all human proteins are tagged with carbohydrate groups that control their activity. Cell-free systems from bacteria lack the machinery to add sugars to proteins, whereas

mammalian cell extracts can affix both desired and undesired chemical modifications. DeLisa and others are developing specific glycosylation modules for bacterial cell-free extracts that would allow researchers to maintain control over the process and keep it precise.

Another obstacle, says Strychalski, is the lack of tools to make cell-free reactions run more like computer code than alchemy. She, for example, is working on refined probes that would allow researchers to follow reactions at the molecular level rather than just indicating success at the end.

Even if these obstacles are overcome, cell-free synthesis will probably never completely replace its cell-based counterpart, Strychalski says. Some companies have already optimized their production facilities to work with cells, and some molecules might always be easier and cheaper to make that way.

For all their apparent simplicity, cell-free systems remain enormously complex. Researchers don't yet have a good enough handle on what's happening in the test tube to model and scale it up effectively. But as cell-free systems take their place in the biomanufacturing toolset, they are providing researchers with a new and increasingly attractive option: just add water.

*Nature* **615**, 551-552 (2023)

*doi: https://doi.org/10.1038/d41586-023-00760-4*

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00760-4

- WHERE I WORK
- 13 March 2023

# Snub-nosed monkeys have taught me important life lessons

Zoologist Zuofu Xiang's research in Asia helps governments to protect the populations and teaches him the value of cooperation.

- Andy Tay



Zuofu Xiang is a zoologist at the Central South University of Forestry and Technology in Changsha, China. Credit: Zuofu Xiang

I have been studying the behaviour and conservation of four species of snub-nosed monkey (*Rhinopithecus* spp.) in China since 2002.

The monkeys form stable, mixed-sex bands of several hundred members, with two to five adult females for every adult male. Some unmated males form groups to improve their chances of reproduction. Young males are expelled from mixed-sex groups to prevent in-breeding.

In this photo, I'm observing a group of unmated male black-and-white snub-nosed monkeys (*R. bieti*) in Baima Xueshan National Nature Reserve in China's Yunnan province. It took about two years for the monkeys to become comfortable with our presence and come that close to us.

My team and I found that unmated males stay separate from but close to mixed-sex bands. In this way, they can mate with females when a resident male lead from a mixed-sex group is temporarily absent.

Studying these monkeys has taught me the importance of cooperation in my personal life and work. Male leaders in mixed-sex groups join forces to defend their positions and prevent hostile takeover by unmated male monkeys. And female monkeys nurse one another's offspring to help infants survive. I encourage my students to do likewise, supporting one another and forming collaborations.

My research has helped me to update the Chinese government on the monkeys' population status and the threats they face, and to suggest ideas for their conservation. My team and I found that infant-monkey deaths increased every year in July and August in Honglaxueshan National Nature Reserve in Tibet. We learnt that that's when farmers were looking for the highly desirable Song Rong mushroom (*Tricholoma matsutake*), and that they would tease the monkeys, causing them to feel threatened and to abandon their infants. Local authorities used our information to tell the villagers to leave the monkeys alone.

*Nature* **615**, 554 (2023)

*doi: https://doi.org/10.1038/d41586-023-00761-3*

---

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00761-3

# Research

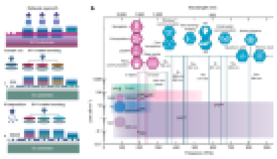
- **Ultrafast tunable lasers using lithium niobate integrated photonics** [15 Mar 2023]
  • A frequency-tunable laser based on a hybrid silicon nitride and lithium niobate integrated photonic platform has a fast tuning rate and could be used for optical ranging applications.
- **Photonically active bowtie nanoassemblies with chirality continuum** [15 Mar 2023]
  • Self-limited assembly of 'imperfect' chiral nanoparticles enables formation of bowtie-shaped microparticles with size monodispersity and continuously variable chirality to be used for printing photonically active metasurfaces.
- **Control of stereogenic oxygen in a helically chiral oxonium ion** [15 Mar 2023]
  • The design, synthesis and characterization of a helically chiral triaryloxonium ion is reported, which is an example of a chiral non-racemic and configurationally stable molecule in which the oxygen atom is the sole stereogenic centre.
- **Regime shift in Arctic Ocean sea ice thickness** [15 Mar 2023]
  • A simple model describes the stochastic process of dynamic sea ice thickening, shows how reduced residence time affects changes in ice thickness and highlights the enduring impact of climate change on the Arctic Ocean.
- **Molecular fate-mapping of serum antibody responses to repeat immunization** [16 Jan 2023]
  • Serum antibody responses to sequential homologous booster vaccines derive overwhelmingly from primary cohort B cells at the expense of de novo responses; this ‘primary addiction’ can be overcome by boosting with variant antigens.
- **Fumarate induces vesicular release of mtDNA to drive innate immunity** [08 Mar 2023]
  • Fumarate metabolism regulates the innate immune response through a mechanism in which high levels of fumarate result in the generation of mitochondrial-derived vesicles and the release of mitochondrial DNA into the cytosol, which activates inflammatory pathways.
- **Structural basis for bacterial energy extraction from atmospheric hydrogen** [08 Mar 2023]

• Structural and biochemical studies of the Mycobacterium smegmatis hydrogenase Huc provides insights into how [NiFe] hydrogenases oxidize trace amounts of atmospheric hydrogen and transfer the electrons liberated via quinone transport.

- ## Common orthopaedic trauma may explain 31,000-year-old remains [15 Mar 2023]
  •

- ## Reply to: Common orthopaedic trauma may explain 31,000-year-old remains [15 Mar 2023]
  •

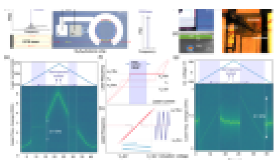
---



## Extending the spectrum of fully integrated photonics to submicrometre wavelengths

28 September 2022

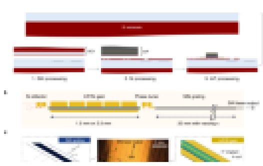
Minh A. Tran, Chong Zhang, … Tin Komljenovic



## Low-noise frequency-agile photonic integrated lasers for coherent ranging

20 June 2022

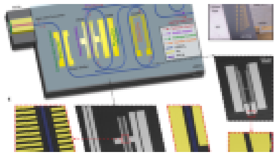
Grigory Lihachev, Johann Riemensberger, … Tobias J. Kippenberg



## High-performance lasers for fully integrated silicon nitride photonics

17 November 2021

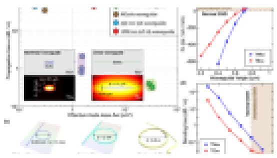
Chao Xiang, Joel Guo, … John E. Bowers



## [Integrated Pockels laser]

12 September 2022

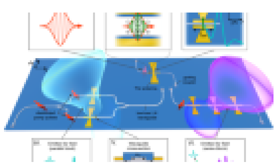
Mingxiao Li, Lin Chang, … Qiang Lin



## [High-yield, wafer-scale fabrication of ultralow-loss, dispersion-engineered silicon nitride photonic circuits]

16 April 2021

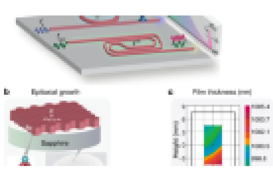
Junqiu Liu, Guanhao Huang, … Tobias J. Kippenberg



## [Terahertz waveform synthesis in integrated thin-film lithium niobate platform]

04 January 2023

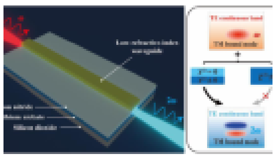
Alexa Herter, Amirhassan Shams-Ansari, … Ileana-Cristina Benea-Chelmus



### Aluminum nitride nanophotonics for beyond-octave soliton microcomb generation and self-referencing

14 September 2021

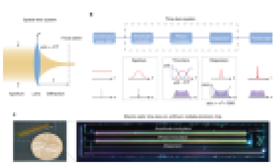
Xianwen Liu, Zheng Gong, … Hong X. Tang



### Efficient second harmonic generation by harnessing bound states in the continuum in semi-nonlinear etchless lithium niobate waveguides

01 November 2022

Xueshi Li, Jiantao Ma, … Jin Liu



### Integrated femtosecond pulse generator on thin-film lithium niobate

16 November 2022

Mengjie Yu, David Barton III, … Marko Lončar

- Article
- Open Access
- Published: 15 March 2023

# Ultrafast tunable lasers using lithium niobate integrated photonics


- Viacheslav Snigirev,
- Annina Riedhauser,
- Grigory Lihachev,
- Mikhail Churaev,
- Johann Riemensberger,
- Rui Ning Wang,
- Anat Siddharth,
- Guanhao Huang,
- Charles Möhl,
- Youri Popoff,
- Ute Drechsler,
- Daniele Caimi,
- Simon Hönl,
- Junqiu Liu,
- Paul Seidler &
- …
- Tobias J. Kippenberg

Show authors



*Nature* **volume 615**, pages 411–417 (2023)



- 4584 Accesses
- 46 Altmetric
- Metrics details


## Abstract

Early works[1] and recent advances in thin-film lithium niobate ($LiNbO_3$) on insulator have enabled low-loss photonic integrated circuits[2,3], modulators with improved half-wave voltage[4,5], electro-optic frequency combs[6] and on-chip electro-optic devices, with applications ranging from microwave photonics to microwave-to-optical quantum interfaces[7]. Although recent advances have demonstrated tunable integrated lasers based on $LiNbO_3$ (refs. [8,9]), the full potential of this platform to demonstrate frequency-agile, narrow-linewidth integrated lasers has not been achieved. Here we report such a laser with a fast tuning rate based on a hybrid silicon nitride ($Si_3N_4$)–$LiNbO_3$ photonic platform and demonstrate its use for coherent laser ranging. Our platform is based on heterogeneous integration of ultralow-loss $Si_3N_4$ photonic integrated circuits with thin-film $LiNbO_3$ through direct bonding at the wafer level, in contrast to previously demonstrated chiplet-level integration[10], featuring low propagation loss of 8.5 decibels per metre, enabling narrow-linewidth lasing (intrinsic linewidth of 3 kilohertz) by self-injection locking to a laser diode. The hybrid mode of the resonator allows electro-optic laser frequency tuning at a speed of $12 \times 10^{15}$ hertz per second with high linearity and low hysteresis while retaining the narrow linewidth. Using a hybrid integrated laser, we perform a proof-of-concept coherent optical ranging (FMCW LiDAR) experiment. Endowing $Si_3N_4$ photonic integrated circuits with $LiNbO_3$ creates a platform that combines the individual advantages of thin-film $LiNbO_3$ with those of $Si_3N_4$, which show precise lithographic control, mature manufacturing and ultralow loss[11,12].

# Main

Lithium niobate ($LiNbO_3$) is an attractive material for electro-optic devices and has been widely used for many decades. It exhibits a wide transparency window from ultraviolet to mid-infrared wavelengths and has a large Pockels coefficient of 32 pm $V^{-1}$, enabling efficient, low-voltage and high-speed modulation. Integrated photonics based on materials exhibiting the Pockels effect—such as aluminium nitride[13]—have been demonstrated before, but only recently for $LiNbO_3$ (ref. [14]). Following the commercial availability of $LiNbO_3$ on insulator via wafer bonding and smart-cut, there

has also been substantial progress in the etching of low-loss $LiNbO_3$ waveguides, culminating in ring resonators with an intrinsic $Q$-factor of 10 $\times 10^6$ (ref. [2]). The majority of these achievements have utilized argon ion beam etching to manufacture partially etched ridge waveguide structures, which enabled modulators operating at complementary metal–oxide–semiconductor (CMOS) voltages[4], quadrature-phase-shift-keying modulators[15] and electro-optic frequency combs[6]. Furthermore, the platform has provided a route to creating interfaces using cavity electro-optics that efficiently couple microwave-to-optical fields[7]. In addition to direct etching, heterogeneous integration of $LiNbO_3$ chiplets onto silicon nitride ($Si_3N_4$)[10] or silicon[16] photonic integrated circuits (PICs) has recently been demonstrated.

Beyond applications for electro-optic modulators, a $LiNbO_3$ integrated photonics platform with a large Pockels coefficient and low propagation loss fulfils all the requirements for realizing integrated narrow-linewidth and frequency-agile laser sources, which feature ultrafast, linear and mode-hop-free tuning. Although integrated lasers have made major advances recently, culminating in hybrid self-injection locked lasers based on high-$Q$ $Si_3N_4$ integrated microresonators that reach fibre laser coherence[17,18], that is, subhertz Lorentzian linewidth, these lasers lack fast frequency actuation. Although integrated narrow-linewidth lasers with similar performance have recently been demonstrated using monolithically integrated piezoelectrical stress-optical actuation that is flat and with megahertz actuation bandwidth[19,20], lasers based on $LiNbO_3$ integrated photonic circuits have the potential for vastly faster tuning, with flat frequency response, at substantially lower drive voltages, and do not exhibit excitations of parasitic vibrational modes of the photonic chip, as in the case of piezoelectrical actuation. An electrically pumped hybrid $LiNbO_3$/III–V laser has been demonstrated using a Vernier-filter-based scheme[8,9], but has not yet achieved this capability. Lasers based on $LiNbO_3$ photonic integrated circuits have the potential to realize a host of laser structures, such as widely tunable Vernier lasers or mode-hop-free lasers for a multitude of applications, including frequency-modulated continuous-wave (FMCW) light detection and ranging (LiDAR)[21], optical coherence tomography,

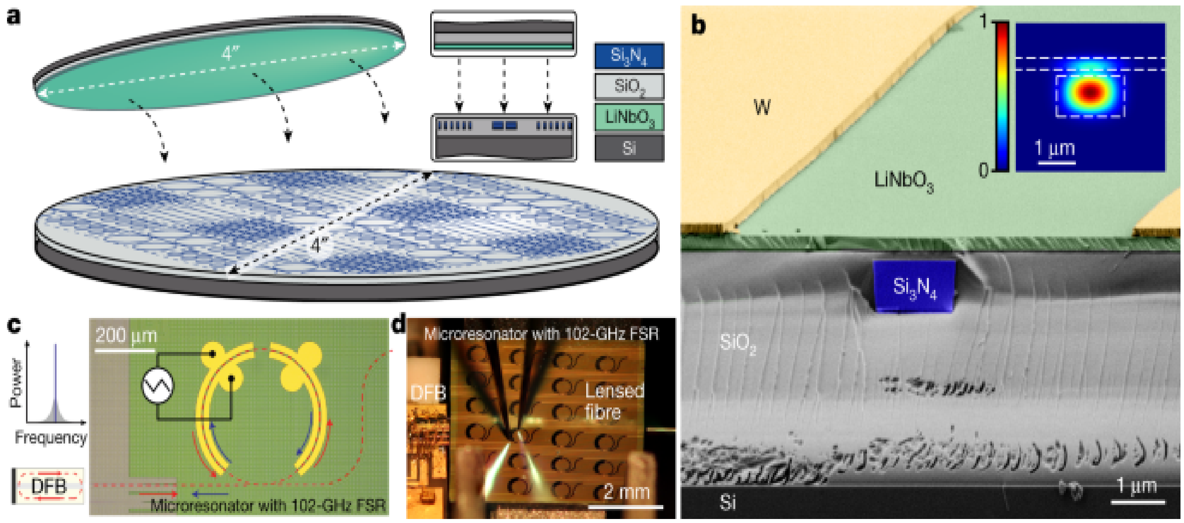
frequency metrology or trace-gas spectroscopy[22], which utilize both frequency agility and narrow linewidth. Here we demonstrate $LiNbO_3$-based integrated lasers that achieve narrow linewidth (kilohertz level) while exhibiting extreme frequency agility, allowing a petahertz-per-second tuning rate. This is achieved on a heterogeneously integrated platform combining ultralow-loss $Si_3N_4$ photonic waveguides[23] with thin-film $LiNbO_3$ by wafer-scale bonding[24]. Our hybrid platform uses a $Si_3N_4$–$LiNbO_3$ chip that is butt-coupled to an indium phosphide (InP) distributed feedback (DFB) diode laser. The $Si_3N_4$ photonic integrated circuits are manufactured using the photonic Damascene process[23] and feature tight optical confinement, ultralow propagation loss (<2 dB $m^{-1}$), low thermal absorption heating and high-power handling. They can be manufactured at the wafer scale with high yield and are already available from a commercial foundry. Additional advantages of the $Si_3N_4$ platform include low gain from the Raman and Brillouin nonlinearities and radiation hardness. This heterogeneous $Si_3N_4$–$LiNbO_3$ platform enables high-$Q$ microresonators with a median intrinsic cavity linewidth of 44 MHz, provides a near-unity yield of bonded devices, and exhibits low, compared with $LiNbO_3$ ridge waveguides, insertion loss of 3.9 dB per facet[24]. In addition, the heterogeneous $Si_3N_4$–$LiNbO_3$ platform does not exhibit bend-induced mode mixing owing to the birefrigence, as is typically the case for $LiNbO_3$ ridge waveguides. Combining the unique properties of both materials into a single heterogeneous integrated platform enables laser self-injection locking with two orders of magnitude of laser frequency noise reduction and a petahertz-per-second frequency tuning rate.

# Heterogeneous integration of LiNbO3 on Si3N4 PIC

Our fabrication method combines the processes of photonic Damascene $Si_3N_4$ waveguide fabrication with wafer-scale bonding[25] to enable electro-optic modulation on passive, ultralow-loss $Si_3N_4$, as schematically depicted in Fig. 1a. Our process starts with the fabrication of a patterned and planarized $Si_3N_4$ substrate using the photonic Damascene process (see

details in Methods). A silicon dioxide ($SiO_2$) interlayer is deposited on the substrate, followed by densification. The interlayer is then polished to reduce the remaining topography and set the desired thickness. A root-mean-square roughness of less than 0.4 nm over an area of a few square micrometres and of only a few nanometres over an area of several hundred micrometres are necessary for bonding[24]. Next, a few-nanometre-thick alumina layer is deposited by atomic layer deposition on both the donor ($LiNbO_3$ on insulator) and the acceptor (planarized $Si_3N_4$ photonic circuit containing 100-mm wafer) wafers before contact bonding and donor-wafer removal. Tungsten electrodes are then manufactured by sputtering and reactive ion etching. At this point, the areas of the coupling facets and the tapered sections of the $Si_3N_4$ waveguides are cleared from $LiNbO_3$ by physical etching with argon ions, so that the laser light can first couple into the chip using inverse tapers[26] before entering the transitioning to the $LiNbO_3$-covered area. Finally, chip release is performed by chip-facet definition by deep $SiO_2$ and silicon etching, followed by chip separation through backside silicon lapping. Figure 1b depicts a scanning electron microscope (SEM) cross-section of the heterogeneously integrated $LiNbO_3$-on-$Si_3N_4$ waveguide with the following layer thicknesses: bottom silica cladding, 4 μm; $Si_3N_4$, 950 nm; silica top cladding, 150 nm; $LiNbO_3$, 300 nm; metal electrodes, 200 nm (the original SEM image is shown in Extended Data Fig. 1). The inset of Fig. 1b shows a simulation of the spatial distribution of the electric-field amplitude in the hybrid mode of our device with 12% participation ratio in $LiNbO_3$. Statistical analysis of the resonator transmission spectra reveals a 44-MHz median intrinsic cavity linewidth corresponding to a $Q$-factor of $4.8 \times 10^6$ and linear propagation loss of 8.5 dB $m^{-1}$ (Extended Data Fig. 8c).

**Fig. 1: Heterogeneous, low-loss, $Si_3N_4$–$LiNbO_3$ photonic integrated platform for fast tunable self-injection-locked lasers.**




**a**, Schematic illustration of the heterogeneous $Si_3N_4$–$LiNbO_3$ platform realized by heterogeneous integration of a 4″ (100 mm) thin-film $LiNbO_3$ wafer onto a 4″ $Si_3N_4$ wafer, with cross-sections of both wafers. **b**, False-colour SEM image of a heterogeneous $Si_3N_4$–$LiNbO_3$ waveguide cross-section. The original SEM image data are shown in Extended Data Fig. 1. Inset: a finite-difference time-domain simulation of the spatial distribution of the hybrid transverse electric mode's electric-field amplitude with 12% participation in $LiNbO_3$, electric-field maximum is coloured in red and minimum in blue. **c**, Schematic illustration of the self-injection locking principle. The optical path is marked with the dashed red line. The red arrow shows the forward optical wave and the blue arrow shows the reflected optical wave from a microresonator. Laser wavelength tuning is achieved by applying a voltage signal (for example, a linear ramp) on the tungsten electrodes. The structures in yellow are the tungsten electrodes. **d**, Photo of the set-up with a DFB laser butt-coupled to a heterogeneous $Si_3N_4$–$LiNbO_3$ chip (sample D67_01b C16 WG 4.2). A pair of probes touch the electrodes for electro-optic modulation, and a lensed fibre collects the output radiation.

# Laser self-injection locking

Laser self-injection locking is initiated by butt-coupling of an InP DFB diode laser to the heterogeneous $Si_3N_4$–$LiNbO_3$ chip (Fig. 1c,d), and tuning the laser current to match the output frequency to the resonance frequency of

the heterogeneous $Si_3N_4$–$LiNbO_3$ microresonator. The optical backreflection on surface or volumetric inhomogeneities inside the microresonator provides spectrally narrowband feedback to the laser diode by coupling clockwise and anticlockwise propagating modes. Light in the clockwise mode radiates back to the laser bearing the power fraction given by the reflection coefficient $R$, which depends on the interaction strength of the modes and the resonator coupling efficiency.

The laser diode is forced to oscillate at the frequency of the cavity resonance in the self-injection-locked regime. Assuming that the frequency noise of the laser is white, the frequency noise suppression ratio[27] is:

$$\frac{{\delta }{\omega }}{{\delta }{{\omega }}_{{\rm{free}}}}\approx \frac{{Q}_{{\rm{DFB}}}^{2}}{{Q}^{2}}\frac{1}{16R(1+{\alpha }_{{\rm{g}}}^{2})},$$
(1)

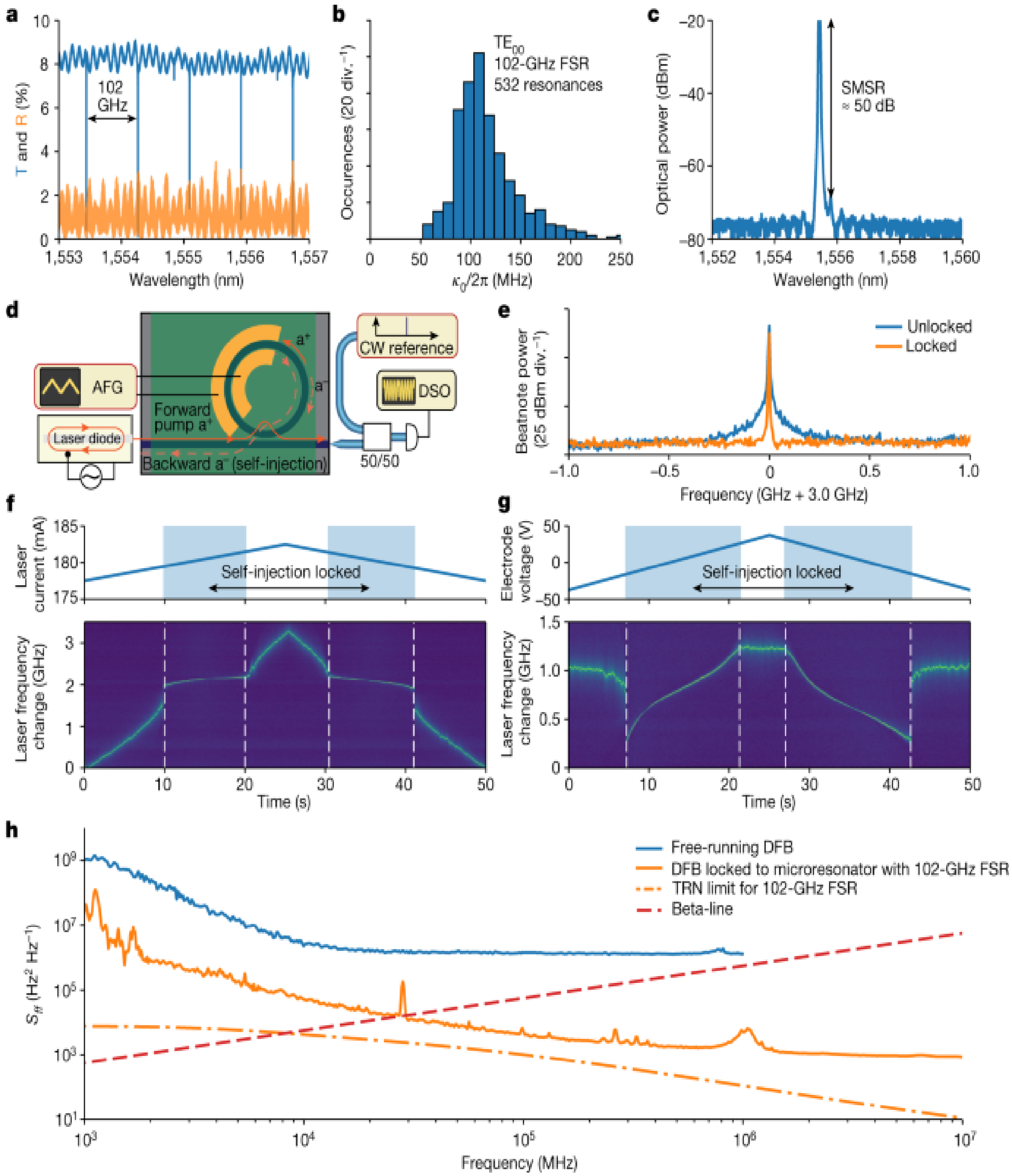
where $\delta\omega_{\rm free}/2\pi$ is the linewidth of the free-running DFB laser; $\delta\omega/2\pi$ is the linewidth of self-injection-locked DFB laser; $Q_{\rm DFB}$ and $Q = \omega/\kappa$ are the quality factors of the laser diode cavity and of the microresonator mode, respectively (with $\kappa = \kappa_{\rm ex} + \kappa_0$, where $\kappa_0$ and $\kappa_{\rm ex}$ are the intrinsic cavity decay rate and bus–waveguide coupling rate, respectively); and $\alpha_{\rm g}$ is the phase-amplitude coupling factor. Self-injection locking occurs within a finite frequency interval around the cavity resonance. The locking bandwidth $\Delta\omega_{\rm lock}$ is given, assuming large intermodal interaction strength and high coupling efficiency, by[27]:

$$\Delta {\omega }_{{\rm{lock}}}\approx \sqrt{R(1+{\alpha }_{{\rm{g}}}^{2})}\frac{\omega }{{Q}_{{\rm{DFB}}}}$$
(2)

To strongly reduce the laser linewidth and increase the frequency locking range, a high-$Q$ resonance and strong reflection are desirable. The device used in our experiments features a 102-GHz free spectral range (FSR) and a resonance total linewidth of $\kappa/2\pi = 100$ MHz (Fig. 2a,b) operated close to critical coupling. The intrinsic loss of the microresonator $\kappa_0/2\pi \approx 50$ MHz indicates a waveguide linear propagation loss of 8.5 dB $m^{-1}$. The power

reflection of the device reaches 3% (see Fig. 2a and Extended Data Fig. 2 for full spectrum) and features both the narrowband reflection ($R$) of the microresonator and the wideband sinusoidal modulation by spurious reflections from the chip facet, as well as the transitions between the inverse tapers and the heterogeneous $Si_3N_4$–$LiNbO_3$ waveguide, which can be mitigated using tapered transitions. Tapered transition in $LiNbO_3$ also decreases insertion losses to 2.5 dB per facet[24]. The reflection from the chip facets can be reduced by using angled output tapers. Despite the weak backreflection contrast (see the characterization of other devices from the wafer in Extended Data Fig. 8), injection locking is observed due to the narrow linewidth of the optical resonance. Laser stabilization for any level of intrinsic backscattering can be further improved by introducing an on-chip drop-port-coupled loop mirror. Adjusting optical feedback by tuning the drop-port mirror coupling and feedback phase allows improvement of the locking range and frequency noise suppression[28]. The self-injection-locked DFB emission spectrum (Fig. 2c) indicates a lasing wavelength of 1,555.4 nm with a side-mode-suppression ratio of 50 dB. To characterize laser self-injection locking, a heterodyne beatnote of the unlocked or the locked DFB laser with a reference laser is generated on a fast photodiode and processed using an electrical spectrum analyser (Fig. 2d). We observe beatnote narrowing upon locking of the DFB laser (Fig. 2e). When varying the laser current of the DFB, we found the regions where there is almost no laser frequency tuning because of the self-injection locking (Fig. 2f). To reveal the locking bandwidth, we set the DFB current inside the locked state and scanned the cavity resonance by applying a triangular voltage chirp to the electrodes (Fig. 2g). The self-injection locking is achieved within a frequency span of around 1 GHz; however, linear tuning is observed only within a 600-MHz band due to the low backreflection of the heterogeneous $Si_3N_4$–$LiNbO_3$ microresonator.

**Fig. 2: Characterizing the integrated hybrid heterogeneous $Si_3N_4$–$LiNbO_3$ platform.**




**a**, Transmission (T, blue) and reflection (R, orange) spectra of a heterogeneous $Si_3N_4$–$LiNbO_3$ 102-GHz-FSR microresonator (see Extended Data Fig. 2 for the full dataset). **b**, The histogram shows the distribution of linewidths of 532 resonances for the fundamental transverse electric mode $TE_{00}$ of the 102-GHz-FSR device with a median linewidth of about 100 MHz, corresponding to a quality factor of $1.9 \times 10^6$ ($\kappa_0$ is the intrinsic cavity

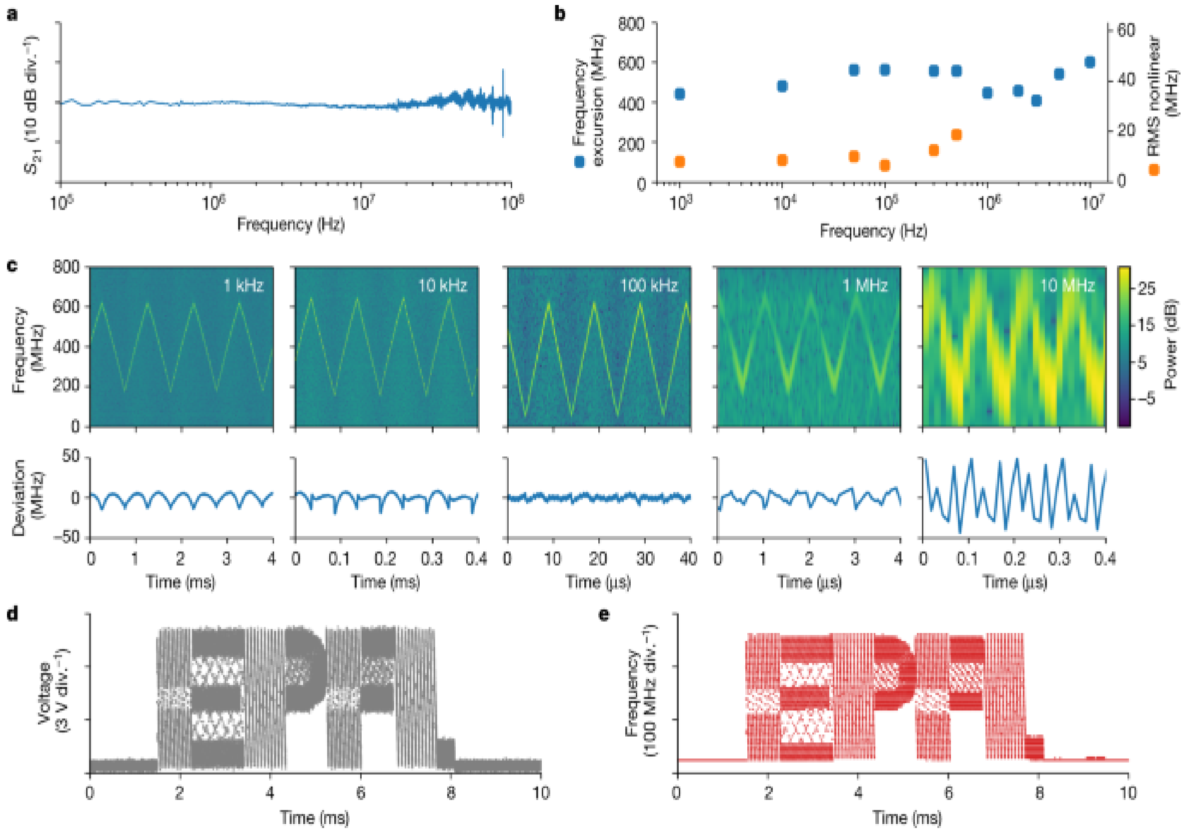
decay rate). **c**, Optical spectrum of the free-running DFB laser diode with 50 dB side mode suppression ratio (SMSR). **d**, Experimental set-up for linewidth measurements with the hybrid integrated laser using the heterodyne beatnote method. AFG, arbitrary function generator; DSO, digital storage oscilloscope. The forward pump wave $a^+$ is marked by a solid red line, and the reflected backward wave $a^-$ by a dashed red line. **e**, Comparison of the laser linewidth for the free-running DFB case and the case where the DFB is self-injection-locked to a heterogeneous $Si_3N_4$–$LiNbO_3$ microresonator. **f**, Time–frequency map of the beatnote showing the laser frequency change on linear modulation of the diode current. The white dashed lines mark the boundaries of the self-injection locking bandwidth, where almost no laser frequency change is observed. **g**, Time–frequency map of the beatnote showing the laser frequency change upon linear tuning of the cavity resonance by applying voltage to the electrodes. The DFB current remained fixed in the self-injection locking range. **h**, Frequency noise spectra of the free-running DFB (blue) and the DFB self-injection-locked to the 102-GHz-FSR heterogeneous $Si_3N_4$–$LiNbO_3$ microresonator (orange). The evaluated thermo-refractive noise (TRN) limit and the beta-line are given for reference (orange dash-dotted and red dashed lines, respectively).

Next, we measured the (single sided) frequency noise spectral density $S_{ff}(f)$ of the DFB diode laser in the free-running and self-injection-locked regimes (see Methods for details and Fig. 2h for results). The laser self-injection locking suppresses the frequency noise by at least 20 dB across all frequency offsets. We find the intersection point of the frequency noise curve and the beta-line[29] at 30 kHz (Fig. 2h). The full-width at half-maximum (FWHM) linewidth, which is calculated by integration of the frequency noise from the beta-line to the inverse integration time, is 56 kHz at 0.1 ms integration time, 262 kHz at 1 ms and 1.1 MHz at 100 ms. The laser frequency noise reaches a horizontal plateau (white noise floor) of $10^3$ $Hz^2$ $Hz^{-1}$ at a 3-MHz offset, which corresponds to an intrinsic laser linewidth of 3.14 kHz.

## Petahertz-per-second frequency-agile laser tuning

To measure the voltage-to-frequency response of the heterogeneous $Si_3N_4$–$LiNbO_3$ microresonator, the signal from a network analyser was applied to the electrodes, and the laser frequency was fixed on the slope of the cavity resonance. This measurement reveals a key advantage of the heterogeneous $Si_3N_4$–$LiNbO_3$ platform—the modulation response function for the 102-GHz-FSR microresonator is flat up to the cavity linewidth of 100 MHz (Fig. 3a). To demonstrate the frequency agility of the laser and the response of the laser frequency to a large-amplitude voltage modulation, the DFB laser was self-injection-locked to the microresonator resonance and a triangular voltage signal of 25 Vpp with modulation frequencies ranging from 1 kHz to 10 MHz was applied. Neither signal pre-distortion nor active feedback was applied to the driving signal. The applied voltage modulates the refractive index of $LiNbO_3$ via the Pockels effect and shifts the cavity resonance, forcing the laser to follow the resonance as long as it remains within the overall locking range. To reveal the time-varying frequency-tuning characteristics for large signal modulation in the self-injection-locked state, the heterodyne beatnote of the hybrid integrated laser with the reference laser was recorded on a fast photodiode. Frequency excursions remained on the level of 500 MHz, independently of the modulation frequency, whereas the nonlinearity tended to increase with increased modulation frequency. The minimum nonlinearity of 1% of the frequency excursion is observed at a tuning rate of 100 kHz with a tuning efficiency of 28 MHz $V^{-1}$. The top row in Fig. 3c shows the processed laser frequency spectrograms, which are calculated by time-segmented Fourier transformation, and the bottom row shows the corresponding residuals after a perfect triangular modulation is fit to the data. Figure 3b shows the laser frequency excursion and the root-mean-squared deviation of the measured profiles from a perfect triangular frequency modulation determined by curve fitting. Additional data on tuning efficiency and hysteresis are shown in Extended Data Figs. 3 and 4. The demonstrated frequency excursion of 600 MHz in 50 ns equates to an ultrafast frequency agility of 12 PHz $s^{-1}$.

**Fig. 3: Electro-optic frequency tuning of the self-injection-locked DFB laser.**




**a**, Measured response of the electro-optic modulation for the heterogeneous $Si_3N_4$–$LiNbO_3$ device utilizing tungsten electrodes. **b**, Frequency excursion (blue) and the absolute root mean square (RMS) deviation of the measured tuning profile from a perfect triangular ramp (orange). The deviation was calculated as the difference between the experimental data and the least-squares fitting. **c**, Top: time–frequency spectrograms of heterodyne beatnote for modulation frequencies from 1 kHz to 10 MHz. Bottom: the deviation of the experimental tuning data from the least-squares fit for the same modulation frequencies. **d**, Voltage profile applied to the electrodes from an arbitrary waveform generator, resembling the EPFL logo. **e**, Measured laser heterodyne beatnote showing laser frequency evolution in the form of the EPFL logo at a tuning rate of 450 Hz $s^{-1}$.

Although highly linear ramp frequency modulation is essential for FMCW LiDAR application, the frequency can be modulated in an arbitrary manner while preserving a high tuning rate. To illustrate this, we programmed an arbitrary waveform generator to reproduce the logo of EPFL (Fig. 3d) and applied the signal to the heterogeneous $Si_3N_4$–$LiNbO_3$ device. The laser

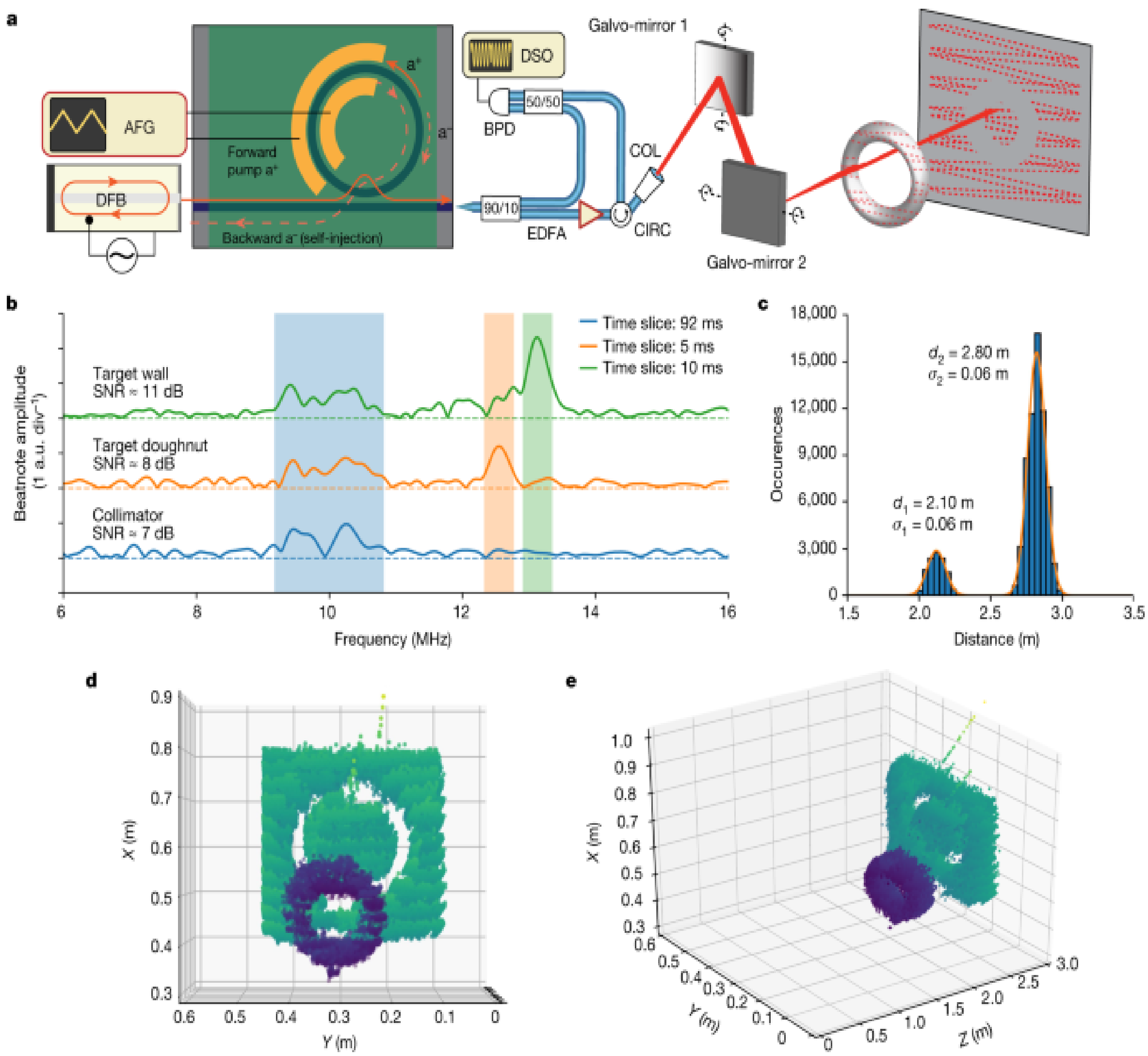
frequency was again determined by heterodyne beatnote with the reference laser, and the result of the time–frequency analysis is depicted in Fig. 3e, showing a tuning rate of 450 THz $s^{-1}$ and a dwell time between points of 200 ns.

# Optical coherent ranging demonstration

To demonstrate the application potential of our laser, we perform a proof-of-concept optical ranging experiment in a lab environment. The FMCW LiDAR method consists of triangular-shaped frequency modulation of the laser source and delayed homodyne detection with the optical signal reflected from the target. Laser phase noise limits the maximum operating distance and ranging precision in this method. However, a key requirement for FMCW LiDAR at long range is frequency agility, that is, to achieve fast, linear and hysteresis-free tuning[30]. The experimental set-up is depicted in Fig. 4a (see Methods for detailed description). The laser beam is scanned across the target scene by means of two galvo-mirrors with triangular driving signals. We used a polystyrene doughnut-like object and a metal sidewall of a rack enclosure as the target. Both objects were located approximately 3 m away from the collimator. A photograph of the target scene and the beam-scanning pattern are depicted in Extended Data Fig. 5. The beatnote between the signal reflected from the target and the local oscillator is detected with a balanced photodiode and recorded by an oscilloscope. We adjust optical polarization with a fibre polarization controller in the reference arm of the delayed self-homodyne set-up to maximize the signal-to-noise ratio (SNR) of the beatnote signal. Zero-padded short-time Fourier transformation is then applied to the collected oscillogram data to retrieve the beatnote spectrum evolution over 128,000 time slices. The time–frequency spectrograms obtained for both the target and the reference Mach–Zehnder interferometer (MZI) are shown in Extended Data Fig. 6. The MZI was used for distance calibration only (we retrieved a resolution of 15 cm), and neither signal pre-distortion nor active feedback was applied. Figure 4b shows three different time frames with beatnotes of the local oscillator with the reflections from the wall, the doughnut and the collimator, and their respective SNR values. Last, the centre frequencies of the beatnote spectra were identified and mapped into the distance domain using the MZI length as a reference. The resulting

distribution of the distance values is plotted as a histogram in Fig. 4c, showing two peaks representing the doughnut at 2.1 m and the wall at 2.8 m. The double-Gaussian fit reveals the statistical distribution of distance values for both objects (Fig. 4c). The point cloud of the three-dimensional optical ranging is inferred from the distance data and the voltage-to-angle conversion of the galvo-mirror controller; it is shown in Fig. 4d,e, where the point colour encodes the distance from the collimator.

**Fig. 4: Coherent FMCW LiDAR demonstration using hybrid integrated lasers.**




**a**, Schematics of the experimental set-up for coherent optical ranging based on frequency-modulated continuous wave (FMCW) LiDAR. The output signal of the tunable laser source with a linear frequency chirp is split into

two channels for delayed homodyne detection. The signal in the first channel is amplified and, by means of mechanical beam-steering, scans the target. The signal in the second channel is mixed with the fraction of the power of the first channel that was scattered by the target. The beatnote power evolution is recorded by an oscilloscope. AFG, arbitrary function generator; DSO, digital storage oscilloscope; EDFA, erbium-doped fibre amplifier; CIRC, optical circulator; BPD, balanced photodiode; COL, collimator; FPC, fibre polarization controller. **b**, Examples of the delayed homodyne beatnote corresponding to signals from the collimator (blue shaded region in all 3 traces), the doughnut (orange shaded region in orange trace) and the wall (green shaded region in green trace) with the respective SNR values. **c**, Histogram showing the distribution of the calculated values of distance to the target. The two peaks correspond to the reflections from the doughnut and the wall. Both peaks are fitted with a double-Gaussian function with fitting parameters, the mean distance ($d$) and standard deviation ($\sigma$), indicated. **d**,**e**, Point-cloud representation of the measured target scene from different viewing angles.

In summary, we have demonstrated a heterogeneous wafer-scale platform for electro-optic photonic integrated circuits that integrates ultralow-loss $Si_3N_4$ waveguides and thin-film $LiNbO_3$. We show optical microresonators with 44-MHz median intrinsic cavity linewidth, corresponding to linear propagation losses of 8.5 dB $m^{-1}$, wideband uniform bus–waveguide coupling, and flat electro-optical frequency actuation response up to 100 MHz. Endowing ultralow-loss $Si_3N_4$ photonic integrated circuits with on-chip $LiNbO_3$ electro-optic modulation enables a hybrid self-injection-locked laser with simultaneously narrow linewidth and fast tuning of 12 PHz $s^{-1}$. This laser allows FMCW optical ranging without the need for signal pre-distortion or active feedback and with a resolution of around 15 cm. Detailed comparison with other photonic integrated tunable lasers based on InP chips is given in Extended Data Table 1 and with other integrated $LiNbO_3$ platforms in Extended Data Table 2. With future improvements in photonic circuit design and fabrication, such as reducing the interlayer $SiO_2$ thickness and optimizing electrode positions, we believe that our platform will form the basis of fast tunable lasers with 10-ns-level switching time, mode-hop-free tuning over tens of gigahertz, and fundamental linewidths below 100 Hz and kilometre-level coherence length. By fully leveraging the high electro-

optic coefficient of $LiNbO_3$, with further improvements in photonic integrated circuit design, these devices can operate with complementary metal–oxide–semiconductor-compatible voltages, or achieve millimetre-scale distance resolution. Beyond integrated lasers, the hybrid platform can also be used to realize other functions, such as photonic microwave and millimetre-wave tracking generators[31], switching networks for photonic computing[32], boson sampling[33] and integrated transceivers. Moreover, the wide transparency window of both $LiNbO_3$ and $Si_3N_4$ allows such frequency agility to be extended to other wavelength ranges, such as the mid-infrared or the visible, providing a platform for fast tunable lasers for applications in trace-gas sensing[34].

# Methods

## Device fabrication

Our process starts with the fabrication of a patterned and planarized $Si_3N_4$ substrate using the photonic Damascene process[23,35]. Deep-ultraviolet stepper lithography is used to pattern waveguides and microresonators on a silicon substrate with 4-μm-thick thermal wet $SiO_2$. The pattern is then dry-etched into the $SiO_2$ layer to form the waveguide preform, followed by a high-temperature reflow of the waveguide preform[36] to reduce surface roughness. Stoichiometric $Si_3N_4$ is deposited by low-pressure chemical vapour deposition on the patterned substrate, filling the preform and forming the waveguide cores. Chemical mechanical polishing is used to remove excess $Si_3N_4$ and to planarize the wafer top surface. Subsequently, the entire substrate is thermally annealed at 1,200 °C to drive out the residual hydrogen contained in the $Si_3N_4$. The $SiO_2$ interlayer is deposited on the $Si_3N_4$ substrate, densified and subsequently polished using chemical mechanical polishing. Before bonding, a few-nanometre-thick alumina layer is deposited on the donor ($LiNbO_3$ on insulator) and the acceptor ($Si_3N_4$) wafers. After that, both wafers are brought in contact and annealed for several hours at 250 °C. The silicon on the backside of the donor wafer is ground, and the residual silicon after grinding is removed with

tetramethylammonium hydroxide wet etching. The thermal $SiO_2$ is wet-etched with buffered hydrofluoric acid. The wafer bonding yield is 100% and we successfully bonded 5 out of 5 wafers during 3 different fabrication runs. A layer of tungsten is sputtered on the $LiNbO_3$ surface and the electrode pattern is transferred in this layer via fluoride-based reactive ion etching. Finally, $LiNbO_3$ is etched to open the chip facet areas to improve the input coupling of the device on the chips by means of argon ion beam etching. The subsequent chip release is performed in three steps: dry etching of chip boundaries in $SiO_2$ with fluorine-based chemistry, further etching of silicon carrier by the Bosch process and backside wafer grinding.

## Laser frequency noise measurements

The free-running DFB performance characterization is shown in Extended Data Fig. 7. We performed heterodyne beatnote spectroscopy[37] beating the reference external-cavity diode laser (Toptica CTL 1550) with the hybrid integrated laser to reveal the frequency noise of the latter. The beatnote of the two signals was detected on a photodiode and its electrical output was then sent to an electrical spectrum analyser (Rohde & Schwarz FSW43). The recorded data for the in-phase and quadrature components of the beatnote were processed by Welch's method[38] to retrieve the single-sided phase noise power spectral density $S_{\phi\phi}$, which was converted to frequency noise $S_{ff}$ using: $S_{ff}=f^2\times S_{\phi\phi}$. To calculate the laser linewidth, we integrate the frequency noise spectra from the intersection of the power spectral density with the beta-line ${S}_{f}(f)=8\,{\rm{ln}}\,2\times f\,/{{\rm{\pi }}}^{2}$ down to the integration time of measurement[29]. The area under the curve $A$ is then recalculated to provide the FWHM measure of the linewidth using: ${\rm{FWHM}}=\sqrt{8\,{\rm{ln}}\,2\times A}$. Because a rigorous definition of the optical linewidth does depend on the integration time of the measurement, we evaluate the FWHM linewidth as 56 kHz at 0.1 ms integration time, 262 kHz at 1 ms and 1.1 MHz at 100 ms. The phase noise of the reference laser is determined by another beatnote measurement with a commercial ultrastable laser (Menlo ORS).

## Thermo-refractive noise simulations

The frequency stability of the 102-GHz-FSR heterogeneous $Si_3N_4$–$LiNbO_3$ device is primarily limited by the material refractive index fluctuations due to the relatively large material temperature fluctuations at the microresonator scale, that is, thermo-refractive noise. To quantify the noise level in our system, we follow the approach based on the fluctuation–dissipation theorem (FDT), described in refs. [39,40,41], that was originally given by Levin and successfully applied to the thermal noise analysis of LIGO’s mirrors. As FDT relates fluctuations of a system to how the system dissipates energy, we simulate the noise levels with the finite-element method by testing how the system dissipates in response to a probe force. As the fractional thermo-refractive noise $\frac{{\delta }\omega }{\omega }=\int {\rm{d}}{\bf{r}}q({\bf{r}}){\delta }T({\bf{r}})$ of our device is a weighted average of the temperature fluctuations $\delta T(\mathbf{r})$ determined by the optical field distribution $\mathbf{e}(\mathbf{r})$ with radius vector $\mathbf{r}$, to find out its magnitude at a particular Fourier frequency $f$, we apply a sinusoidal entropy oscillation (energy-conjugated with temperature) at this frequency, with the same weight $\mathbf{q}(\mathbf{r})$ mimicking the field distribution, to our system in the simulation. The corresponding power dissipated $W_{\mathrm{diss}}$ in the system is retrieved from the simulation and is used to calculate the thermo-refractive noise power spectral density ${S}_{\frac{{\delta }\omega }{\omega }}(\,f)$ at this particular frequency using FDT. The device field distribution and the heat propagation simulated in the described steps are performed on COMSOL Multiphysics.

## Coherent ranging experiment

The laser diode is edge-coupled to the heterogeneous $Si_3N_4$–$LiNbO_3$ chip with 200-nm-thick tungsten electrodes deposited along the $Si_3N_4$ waveguide on $LiNbO_3$. The laser frequency tuning is achieved by locking the laser to a cavity resonance, fixing the DFB current, and tuning the cavity resonance via Pockels effect by a voltage applied to the electrodes. The triangular ramp signal from the arbitrary waveform generator with 0.5 Vpp amplitude and 100-kHz frequency is further amplified up to 25 Vpp by a high-voltage amplifier (Falco Systems) with 5-MHz bandwidth. No additional pre- or post-processing (linearization) was utilized for the laser frequency ramp for the coherent ranging experiment. We used the cavity resonance

corresponding to 179-mA DFB current. To calibrate the frequency excursion, the 5% fraction of the optical signal was sent to a reference MZI fibre interferometer. The MZI optical length of 13.18 m was found by an independent measurement involving a tunable diode laser scan calibrated by a frequency comb. Taking the measured MZI optical length and beatnote frequency values, a distance resolution of 15 cm is inferred. Ninety-five per cent of light is split into two paths: the local oscillator path (10%) and the target path (90%). The signal in the target path is amplified by an erbium-doped fibre amplifier (Calmar) from 150 μW up to 4 mW and directed to the collimator with the 8-mm aperture set to match the target distance range of 3 m. We use the galvo scanner (Thorlabs GVS112) for the beam-steering. Two mirrors were controlled by linear ramp signals of 3 Hz and 60 Hz rates with the amplitude and offset values chosen to ensure that the scanning pattern fully covers the target scene. The data for the point cloud were collected within the total time interval of 1.3 s. The frame rate was limited by the galvo scanning speed and the Doppler broadening that is imparted by the rapidly tilting mirrors.

## Scene reconstruction and signal data processing

The data collected in the FMCW LiDAR experiment were subject to digital signal processing steps to locate the scene elements in space. First, the zero-padded short-time Fourier transforms of the beatnote oscillograms from the target and the reference MZI were evaluated. The Blackman–Harris window function was used with the window size set to one period of the frequency-modulated signal. Second, the obtained time–frequency maps were used to search at any given time frame the frequency values corresponding to the beatnote peak. This set was filtered so that only the data points with beatnote amplitudes above some threshold were considered for further analysis. We then subtract the distance from the laser to the collimator so that the point-cloud distance is given with respect to the collimator aperture position. Finally, the frequency data were converted to the distance domain, using MZI length as a reference, and the Cartesian components of each point were computed from the voltage profile applied to the galvo-mirrors.

## Electro-optic $S_{21}$ response measurement

A continuous-wave (CW) laser at 1,550 nm of 300 μW power from an external-cavity diode laser (Toptica CTL 1550) is coupled into the device using a lensed fibre[26]. The input laser is biased at the slope of the optical resonance. A radiofrequency electrical signal of −5 dBm power is applied from port 1 of the network analyser to the electrodes of the device, and the light intensity modulation is detected by a 12-GHz photodiode (New Focus 1544), which is sent back to port 2 of the network analyser.

### Performance comparison

Performance comparison of tunable laser systems[9,17,20,42,43,44,45,46,47,48,49,50,51,52,53,54] is presented in Extended Data Table 1. The table compares different tunable laser systems in terms of the frequency tuning range, tuning rate, linearity, optical output power and white frequency noise floor. Performance comparison of different integrated $LiNbO_3$-based platforms[2,55,56,57,58,59,60,61,62] is presented in Extended Data Table 2.

## Data availability

The data used to produce the plots within this paper are available at https://doi.org/10.5281/zenodo.7371066.

## Code availability

The code used to produce the plots within this paper is available at https://doi.org/10.5281/zenodo.7371066.

## References


1. Guarino, A., Poberaj, G., Rezzonico, D., Degl'Innocenti, R. & Günter, P. Electro-optically tunable microring resonators in lithium niobate. *Nat. Photon.* **1**, 407–410 (2007).

2. Zhang, M., Wang, C., Cheng, R., Shams-Ansari, A. & Lončar, M. Monolithic ultra-high-*Q* lithium niobate microring resonator. *Optica* **4**,

1536–1537 (2017).

3. Krasnokutska, I., Tambasco, J.-L. J., Li, X. & Peruzzo, A. Ultra-low loss photonic circuits in lithium niobate on insulator. *Opt. Express* **26**, 897–904 (2018).

4. Wang, C. et al. Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages. *Nature* **562**, 101–104 (2018).

5. He, M. et al. High-performance hybrid silicon and lithium niobate mach–zehnder modulators for 100 gbit s-1 and beyond. *Nat. Photon.* **13**, 359–364 (2019).

6. Zhang, M. et al. Broadband electro-optic frequency comb generation in a lithium niobate microring resonator. *Nature* **568**, 373–377 (2019).

7. Lambert, N. J., Rueda, A., Sedlmeir, F. & Schwefel, H. G. L. Coherent conversion between microwave and optical photons—an overview of physical implementations. *Adv. Quantum Technol.* **3**, 1900077 (2020).

8. Han, Y. et al. Electrically pumped widely tunable O-band hybrid lithium niobite/III–V laser. *Opt. Lett.* **46**, 5413–5416 (2021).

9. Op de Beeck, C. et al. III–V-on-lithium niobate amplifiers and lasers. *Optica* **8**, 1288–1289 (2021).

10. Chang, L. et al. Heterogeneous integration of lithium niobate and silicon nitride waveguides for wafer-scale photonic integrated circuits on silicon. *Opt. Lett.* **42**, 803–806 (2017).

11. Puckett, M. W. et al. 422 million intrinsic quality factor planar integrated all-waveguide resonator with sub-mHz linewidth. *Nat. Commun.* **12**, 934 (2021).

12. Liu, K. et al. Ultralow 0.034 db/m loss wafer-scale integrated photonics realizing 720 million $Q$ and 380 μW threshold Brillouin lasing. *Opt. Lett.* **47**, 1855–1858 (2022).

13. Xiong, C., Pernice, W. H. P. & Tang, H. X. Low-loss, silicon integrated, aluminum nitride photonic circuits and their use for electro-optic signal processing. *Nano Lett.* **12**, 3562–3568 (2012).

14. Wang, C. et al. Integrated high quality factor lithium niobate microdisk resonators. *Opt. Express* **22**, 30924–30933 (2014).

15. Xu, M. et al. High-performance coherent optical modulators based on thin-film lithium niobate platform. *Nat. Commun.* **11**, 3911 (2020).

16. Weigel, P. O. et al. Lightwave circuits in lithium niobate through hybrid waveguides with silicon photonics. *Sci. Rep.* **6**, 22301 (2016).

17. Jin, W. et al. Hertz-linewidth semiconductor lasers using cmos-ready ultra-high-*Q* microresonators. *Nat. Photon.* **15**, 346–353 (2021).

18. Li, B. et al. Reaching fiber-laser coherence in integrated photonics. *Opt. Lett.* **46**, 5201–5204 (2021).

19. Liu, J. et al. Monolithic piezoelectric control of soliton microcombs. *Nature* **583**, 385–390 (2020).

20. Lihachev, G. et al. Low-noise frequency-agile photonic integrated lasers for coherent ranging. *Nat. Commun.* **13**, 3522 (2022).

21. Martin, A. et al. Photonic integrated circuit-based FMCW coherent lidar. *J. Lightw. Technol.* **36**, 4640–4645 (2018).

22. Debecker, I., Mohamed, A. K. & Romanini, D. High-speed cavity ringdown spectroscopy with increased spectral resolution by simultaneous laser and cavity tuning. *Opt. Express* **13**, 2906–2915 (2005).

23. Liu, J. et al. High-yield, wafer-scale fabrication of ultralow-loss, dispersion-engineered silicon nitride photonic circuits. *Nat. Commun.* **12**, 2236 (2021).

24. Churaev, M. et al. A heterogeneously integrated lithium niobate-on-silicon nitride photonic platform. Preprint at

https://arxiv.org/abs/2112.02018 (2021).

25. Komljenovic, T. et al. Heterogeneous silicon photonic integrated circuits. *J. Lightw. Technol.* **34**, 20–35 (2016).

26. Liu, J. et al. Double inverse nanotapers for efficient light coupling to integrated photonic devices. *Opt. Lett.* **43**, 3200–3203 (2018).

27. Kondratiev, N. M. et al. Self-injection locking of a laser diode to a high-*Q* WGM microresonator. *Opt. Express* **25**, 28167–28178 (2017).

28. Galiev, R. R., Kondratiev, N. M., Lobanov, V. E., Matsko, A. B. & Bilenko, I. A. Mirror-assisted self-injection locking of a laser to a whispering-gallery-mode microresonator. *Phys. Rev. Appl.* **16**, 064043 (2021).

29. Domenico, G. D., Schilt, S. & Thomann, P. Simple approach to the relation between laser frequency noise and laser line shape. *Appl. Opt.* **49**, 4801–4807 (2010).

30. Behroozpour, B. et al. Electronic-photonic integrated circuit for 3D microimaging. *IEEE J. Solid State Circuits* **52**, 161–172 (2017).

31. Chao, R.-L. et al. Fully integrated photonic millimeter-wave tracking generators on the heterogeneous III–V/Si platform. *IEEE Photon. Technol. Lett.* **30**, 919–922 (2018).

32. Drake, B. L., Bocker, R. P., Lasher, M. E., Patterson, R. H. & Miceli, W. J. Photonic computing using the modified signed-digit number representation. *Opt. Eng.* **25**, 38–43 (1986).

33. Spring, J. B. et al. Boson sampling on a photonic chip. *Science* **339**, 798–801 (2013).

34. Truong, G.-W. et al. Frequency-agile, rapid scanning spectroscopy. *Nat. Photon.* **7**, 532–534 (2013).

35. Pfeiffer, M. H. P. et al. Photonic damascene process for low-loss, high-confinement silicon nitride waveguides. *IEEE J. Sel. Top. Quantum*

*Electron.* **24**, 1–11 (2018).

36. Pfeiffer, M. H. P. et al. Ultra-smooth silicon nitride waveguides based on the Damascene reflow process: fabrication and loss origins. *Optica* **5**, 884–892 (2018).

37. Duthel, T. et al. Laser linewidth estimation by means of coherent detection. *IEEE Photon. Technol. Lett.* **21**, 1568–1570 (2009).

38. Welch, P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **15**, 70–73 (1967).

39. Levin, Y. Internal thermal noise in the LIGO test masses: a direct approach. *Phys. Rev. D* **57**, 659–663 (1998).

40. Kondratiev, N. & Gorodetsky, M. Thermorefractive noise in whispering gallery mode microresonators: analytical results and numerical simulation. *Phys. Lett. A* **382**, 2265–2268 (2018).

41. Huang, G. et al. Thermorefractive noise in silicon-nitride microresonators. *Phys. Rev. A* **99**, 061801 (2019).

42. Fan, Y. et al. Hybrid integrated InP–$Si_3N_4$ diode laser with a 40-Hz intrinsic linewidth. *Opt. Express* **28**, 21713–21728 (2020).

43. Tran, M. A. et al. Ring-resonator based widely-tunable narrow-linewidth Si/InP integrated lasers. *IEEE J. Sel. Top. Quantum Electron.* **26**, 1–14 (2020).

44. Malik, A. et al. Widely tunable, heterogeneously integrated quantum-dot O-band lasers on silicon. *Photon. Res.* **8**, 1551–1557 (2020).

45. Wan, Y. et al. High speed evanescent quantum-dot lasers on Si. *Laser Photon. Rev.* **15**, 2100057 (2021).

46. Poulton, C. V. et al. Coherent solid-state lidar with silicon photonic optical phased arrays. *Opt. Lett.* **42**, 4091–4094 (2017).

47. Satyan, N., Vasilyev, A., Rakuljic, G., Leyva, V. & Yariv, A. Precise control of broadband frequency chirps using optoelectronic feedback. *Opt. Express* **17**, 15991–15999 (2009).

48. Zhang, X., Pouls, J. & Wu, M. C. Laser frequency sweep linearization by iterative learning pre-distortion for FMCW lidar. *Opt. Express* **27**, 9965–9974 (2019).

49. Okano, M. & Chong, C. Swept source lidar: simultaneous FMCW ranging and nonmechanical beam steering with a wideband swept source. *Opt. Express* **28**, 23898–23915 (2020).

50. DiLazaro, T. & Nehmetallah, G. Large-volume, low-cost, high-precision FMCW tomography using stitched DFBS. *Opt. Express* **26**, 2891–2904 (2018).

51. Han, Y. et al. Electrically pumped widely tunable o-band hybrid lithium niobite/III–V laser. *Opt. Lett.* **46**, 5413–5416 (2021).

52. Shams-Ansari, A. et al. Electrically pumped laser transmitter integrated on thin-film lithium niobate. *Optica* **9**, 408–411 (2022).

53. Werner, C. S., Yoshiki, W., Herr, S. J., Breunig, I. & Buse, K. Geometric tuning: spectroscopy using whispering-gallery resonator frequency-synthesizers. *Optica* **4**, 1205–1208 (2017).

54. Minet, Y. et al. Pockels-effect-based adiabatic frequency conversion in ultrahigh-*Q* microresonators. *Opt. Express* **28**, 2939–2947 (2020).

55. Boynton, N. et al. A heterogeneously integrated silicon photonic/lithium niobate travelling wave electro-optic modulator. *Opt. Express* **28**, 1868–1884 (2020).

56. Zhang, M. et al. Electronically programmable photonic molecule. *Nat. Photon.* **13**, 36–40 (2019).

57. Chang, L. et al. Heterogeneous integration of lithium niobate and silicon nitride waveguides for wafer-scale photonic integrated circuits on silicon. *Opt. Lett.* **42**, 803–806 (2017).

58. Luke, K. et al. Wafer-scale low-loss lithium niobate photonic integrated circuits. *Opt. Express* **28**, 24452–24458 (2020).

59. Ahmed, A. N. R., Shi, S., Mercante, A. J. & Prather, D. W. High-performance racetrack resonator in silicon nitride–thin film lithium niobate hybrid platform. *Opt. Express* **27**, 30741–30751 (2019).

60. Rao, A. et al. High-performance and linear thin-film lithium niobate Mach–Zehnder modulators on silicon up to 50 GHz. *Opt. Lett.* **41**, 5700–5703 (2016).

61. Jin, S., Xu, L., Zhang, H. & Li, Y. $LiNbO_3$ thin-film modulators using silicon nitride surface ridge waveguides. *IEEE Photon. Technol. Lett.* **28**, 736–739 (2016).

62. Ghosh, S., Yegnanarayanan, S., Ricci, M., Kharas, D. & Juodawlkis, P. Low-loss thin film lithium niobate bonded on silicon nitride waveguides. In *Conference on Lasers and Electro-Optics: OSA Technical Digest* STh1F.1 (Optica, 2020); http://opg.optica.org/abstract.cfm?URI=CLEO_SI-2020-STh1F.1

## Acknowledgements

This work was supported by funding from the European Union H2020 Research and Innovation Program under the Marie Sklodowska-Curie grant agreement number 812818 (MICROCOMB) and 722923 (OMT), under the FET-Proactive grant agreement number 732894 (HOT) and grant agreement number 847471 (QUSTEC). It was also supported by funding from the Swiss National Science Foundation under grant agreement number 186364 (QuantEOM) and 201923 (AMBIZIONE), as well as by the Air Force Office of Scientific Research (AFOSR) under award number FA9550-19-1-0250 and by Contract HR0011-20-2-0046 (NOVEL) from the Defence Advanced Research Projects Agency (DARPA). The samples were fabricated in the EPFL centre of MicroNanoTechnology (CMi) and the Binnig and Rohrer Nanotechnology Center (BRNC) at IBM Research. We thank the Cleanroom Operations Team of the BRNC, especially D. Davila Pineda and R. Grundbacher for their help and support.

# Funding

Open access funding provided by EPFL Lausanne.

# Author information

Author notes

1. These authors contributed equally: Viacheslav Snigirev, Annina Riedhauser, Grigory Lihachev, Mikhail Churaev

## Authors and Affiliations

1. Institute of Physics, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland

   Viacheslav Snigirev, Grigory Lihachev, Mikhail Churaev, Johann Riemensberger, Rui Ning Wang, Anat Siddharth, Guanhao Huang, Junqiu Liu & Tobias J. Kippenberg

2. Center for Quantum Science and Engineering, EPFL, Lausanne, Switzerland

   Viacheslav Snigirev, Grigory Lihachev, Mikhail Churaev, Johann Riemensberger, Rui Ning Wang, Anat Siddharth, Guanhao Huang, Junqiu Liu & Tobias J. Kippenberg

3. IBM Research - Europe, Zurich, Ruschlikon, Switzerland

   Annina Riedhauser, Charles Möhl, Youri Popoff, Ute Drechsler, Daniele Caimi, Simon Hönl & Paul Seidler

4. Deep Light SA https://deeplight.pro/

   Johann Riemensberger

5. Integrated Systems Laboratory, Swiss Federal Institute of Technology Zurich (ETH Zürich), Zurich, Switzerland

Youri Popoff

### Contributions

M.C. designed the lithography masks and performed PIC simulations. V.S. and G.L. performed experiments with the help of J.R., M.C. and A.S. R.N.W., A.R., C.M. and J.L. developed the processes and fabricated the samples with assistance from S.H. U.D., Y.P., A.R., R.N.W. and J.L. performed the chemical mechanical polishing for bonding. D.C. performed the wafer bonding. V.S., G.L. and J.R. analysed the data. V.S. and G.H. performed thermo-refractive noise limit simulations. V.S., G.L., J.R. and T.J.K. wrote the manuscript with input from A.R., A.S., J.L. and P.S. P.S. and T.J.K. supervised the project.

### Corresponding authors

Correspondence to Paul Seidler or Tobias J. Kippenberg.

## Ethics declarations

### Competing interests

T.J.K. is a co-founder and shareholder of LiGenTec SA, a foundry commercializing $Si_3N_4$ photonic integrated circuits, as well as DEEPLIGHT SA, a start-up commercializing $Si_3N_4$ photonic integrated circuits-based frequency-agile, low-noise lasers.

## Peer review

### Peer review information

*Nature* thanks Antonella Bogoni and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Scanning electron microscopy image of an heterogeneous $Si_3N_4$-$LiNbO_3$ waveguide.

Original, unprocessed SEM data used to prepare Fig. 1b of the main text.

### Extended Data Fig. 2 Transmission and reflection of a 102-GHz FSR heterogeneous $Si_3N_4$-$LiNbO_3$ microring resonator.

**(a)** Transmission spectrum. **(b)** Reflection spectrum. Data from the sample D67_01b F2 C16 4.3.

### Extended Data Fig. 3 Hysteresis evaluation for the DFB laser diode frequency tuning.

The first row **(a)** represents the heterodyne beatnote evolution for the chirping rate of 1 kHz, 10 kHz, 100 kHz and 1 MHz, and its fit with a perfect triangular ramp. The second row **(b)** is the same data but mirrored with respect to a horizontal axis of 0 MHz. In the third row **(c)**, the data is shifted by half a period to the left, such as the up-ramp in the first row becomes the down-ramp in the third and vice versa. In the last row **(d)**, adding the data patterns of the first and the third row and subtracting the mean value from the sum, one observes the hysteresis-induced deviations between the up-ramp and down-ramp.

### Extended Data Fig. 4 Tuning efficiency of heterogeneous $Si_3N_4$-$LiNbO_3$ samples.

**(a)** Applying a triangular ramp voltage waveform to heterogeneous $Si_3N_4$-$LiNbO_3$ device's electrodes with selected values of the modulation frequency (10 kHz, 100 kHz and 1 MHz) and gradually increasing the signal peak-to-peak amplitude, one can observe linear growth of the DFB frequency excursion. **(b)** To retrieve the tuning efficiency values, the linear model fit of the data can be performed in the voltages range where the excursion is smaller than the locking bandwidth limit of ~1 GHz (see inset).

## Extended Data Fig. 5 FMCW LiDAR set-up.

**(a)** The photo of the set-up for coherent optical ranging experiment containing the target - a polystyrene doughnut mounted on the stage and an instrument box wall behind, and the scanning pattern of the galvo-mirrors. **(b,c)** The voltage signal profiles applied to two galvo-mirrors enabling two angular degrees of freedom - $\phi$ and $\theta$ - for scanning, and their fits with a perfect triangular ramp. **(d)** The actual data of the scanning pattern and its reconstruction after fitting the angular coordinates.

## Extended Data Fig. 6 Short-time Fourier transform of the delayed homodyne beatnote oscillogram.

**(a)** Time–frequency map calculated for the target response. **(b)** Time–frequency map for the reference Mach–Zehnder interferometer.

## Extended Data Fig. 7 Free-running DFB characterization.

**(a)** Laser diode optical power in free-space vs. diode current. **(b)** Free-running DFB optical spectra at different driving currents.

## Extended Data Fig. 8 Heterogeneous $Si_3N_4$-$LiNbO_3$ wafer characterization.

**(a,d,g,j)** Transmission of bus–waveguide coupled to heterogeneous $Si_3N_4$-$LiNbO_3$ microresonator with free spectral range (FSR) 102 GHz (C11) from 3 fields of the wafer and different microresonators on a chip (F7 WG3.3, F1 WG3.3, F7 WG4.2, F5 WG4.1); **(b,e,h,k)** Frequency-dependent intrinsic

microresonator loss $\kappa_0/2\pi$ (green) and bus–waveguide coupling $\kappa_{ex}/2\pi$ (blue). **(c,f,i,l)** Histogram of intrinsic microresonator loss rate of wafer D67_01b.

**Extended Data Table 1 Performance comparison of tunable laser systems**
**Extended Data Table 2 Performance comparison of integrated lithium-niobate-based platforms**

# Supplementary information

**Peer Review File**

# Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

# About this article

## Cite this article

Snigirev, V., Riedhauser, A., Lihachev, G. *et al.* Ultrafast tunable lasers using lithium niobate integrated photonics. *Nature* **615**, 411–417 (2023).

https://doi.org/10.1038/s41586-023-05724-2

- Received: 08 October 2021
- Accepted: 11 January 2023
- Published: 15 March 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05724-2

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

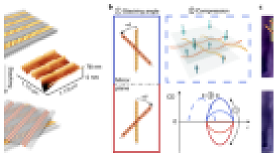
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

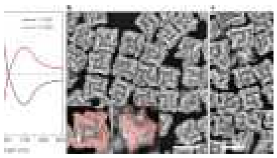
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05724-2



## Mechano-tunable chiral metasurfaces via colloidal assembly

29 April 2021

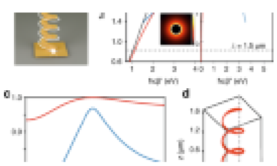
Patrick T. Probst, Martin Mayer, … Andreas Fery



## Amino-acid- and peptide-directed synthesis of chiral plasmonic gold nanoparticles

18 April 2018

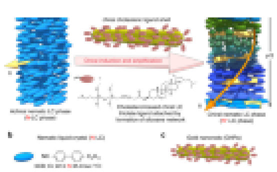
Hye-Eun Lee, Hyo-Yong Ahn, … Ki Tae Nam



## Subwavelength polarization optics via individual and coupled helical traveling-wave nanoantennas

28 August 2019

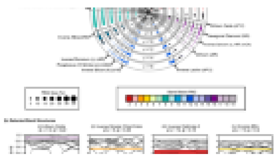
Mengjia Wang, Roland Salut, … Thierry Grosjean



### Chirality amplification by desymmetrization of chiral ligand-capped nanoparticles to nanorods quantified in soft condensed matter

25 September 2018

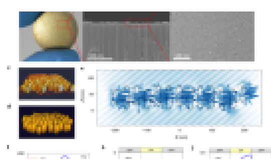
Ahlam Nemati, Sasan Shadpour, … Torsten Hegmann



### The diversity of three-dimensional photonic crystals

05 May 2021

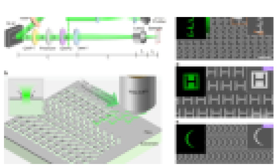
Rose K. Cersonsky, James Antonaglia, … Sharon C. Glotzer



### Polarization-sensitive optoionic membranes from chiral plasmonic nanoparticles

14 March 2022

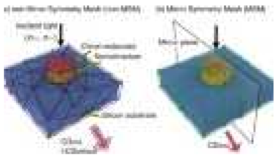
Jiarong Cai, Wei Zhang, … Hua Kuang



### Sub-wavelength patterned pulse laser lithography for efficient fabrication of large-area metasurfaces

03 October 2022

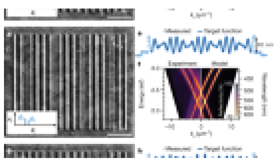
Lingyu Huang, Kang Xu, … Shaolin Xu



[Robust numerical evaluation of circular dichroism from chiral medium/nanostructure coupled systems using the finite-element method]

30 May 2018

Seojoo Lee, Ji-Hun Kang, … Q-Han Park



[Optical Fourier surfaces]

24 June 2020

Nolan Lassaline, Raphael Brechbühler, … David J. Norris

- Article
- Published: 15 March 2023

# Photonically active bowtie nanoassemblies with chirality continuum

- Prashant Kumar,
- Thi Vo,
- Minjeong Cha,
- Anastasia Visheratina,
- Ji-Young Kim,
- Wenqian Xu,
- Jonathan Schwartz,
- Alexander Simon,
- Daniel Katz,
- Valentin Paul Nicu,
- Emanuele Marino,
- Won Jin Choi,
- Michael Veksler,
- Si Chen,
- Christopher Murray,
- Robert Hovden,
- Sharon Glotzer &
- …
- Nicholas A. Kotov

Show authors


*Nature* **volume 615**, pages 418–424 (2023)


- 2870 Accesses

- 69 Altmetric

- Metrics details


## Abstract

Chirality is a geometrical property described by continuous mathematical functions[1,2,3,4,5]. However, in chemical disciplines, chirality is often treated as a binary left or right characteristic of molecules rather than a continuity of chiral shapes. Although they are theoretically possible, a family of stable chemical structures with similar shapes and progressively tuneable chirality

is yet unknown. Here we show that nanostructured microparticles with an anisotropic bowtie shape display chirality continuum and can be made with widely tuneable twist angle, pitch, width, thickness and length. The self-limited assembly of the bowties enables high synthetic reproducibility, size monodispersity and computational predictability of their geometries for different assembly conditions[6]. The bowtie nanoassemblies show several strong circular dichroism peaks originating from absorptive and scattering phenomena. Unlike classical chiral molecules, these particles show a continuum of chirality measures[2] that correlate exponentially with the spectral positions of the circular dichroism peaks. Bowtie particles with variable polarization rotation were used to print photonically active metasurfaces with spectrally tuneable positive or negative polarization signatures for light detection and ranging (LIDAR) devices.

# Main

Mathematical definitions of mirror asymmetry[1,2,3,4,5] recognize the continuity of chiral geometries that can be visualized by stretching macroscale helical springs to obtain coils of different lengths, that is, pitch. At smaller scales, continuously variable chirality can be observed for origami/kirigami sheets[7,8,9], nanocomposites[10,11] and polymeric solids whose shape and circular dichroism (CD) can be varied by external fields[12]. However, chirality in chemistry commonly manifests as a binary property: chiral molecules are either right- or left-handed and the descriptors for stereochemical configurations of enantiomers at molecular scale are correspondingly binary, for example d/l, R/S, M/P and Δ/Λ. The binary chirality of amino acids originates from a high-energy penalty for distorting the optical centre based on *sp*$^3$-carbon atoms. The discreteness of chirality in liquid crystals[13], macromolecular compounds[14], helical polymers[15] and nanoparticles (NPs)[16,17] manifests in abrupt transitions between chiral phases with different crystallinity or particle shapes[18]. Energy penalties are relaxed for large flexible molecules[19], supramolecular complexes[16,20,21] and biopolymers[18,22,23], but the restrictions on chiral shapes remain stringent. Unlike macroscale springs, the helical pitch across the variety of biomolecules varies little. For DNA, protein α-helices and β-sheets, pitch

changes only in the ranges of 11–46, 2.3–5.5 and 7–8 Å, respectively; these narrow ranges are essential for the precise folding of biomolecules[24].

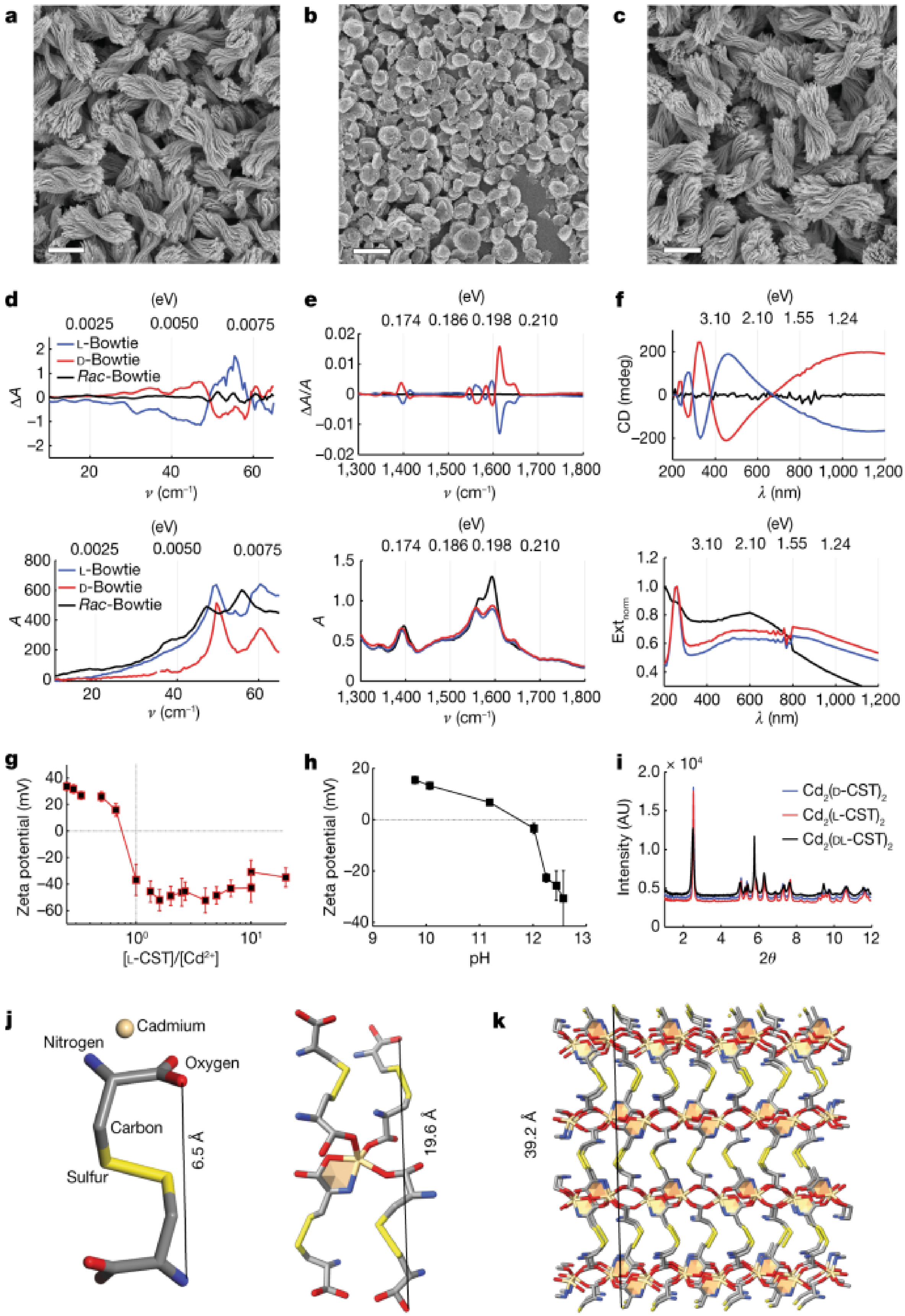
The transition from discrete chiral phases and shapes to a palette of chemical compounds with continuously tuneable chirality would be transformative for the development of chiral photonics, chiral metamaterials, biochemical separations and chiral catalysis. The availability of continuously variable chiral compounds is essential for establishing fundamental correlations between chirality measures and chemical properties. For example, the attempts to correlate optical activity with a variety of chirality measures have largely failed[25,26], but they could be possible for chiral nanostructures and their assemblies due to the differences in physics of chiroptical activity in them compared to molecules with binary chirality

Continuously variable chiral geometries become possible for nanostructured microparticles with bowtie shapes. They are hierarchically assembled from nanoribbons containing helical chains of cystine (CST, dipeptide of cysteine through the S–S bridge) interconnected by $Cd^{2+}$ ions. The balance between short- and long-range interactions and defect tolerance of the electrostatically restricted assembly process[27] enables the synthesis of bowties with widely tuneable pitch, width, thickness and length.

## Bowtie synthesis and structure

Mixing an aqueous solutions of $Cd^{2+}$ and l- or d-cystine (Supplementary Fig. 1a,b) results in bowties (Methods) with pitch (*p*), width (*w*), thickness (*t*) and length (*l*) of 4.1, 1.3, 0.5 and 3.1 µm, respectively (Fig. 1a). The process is 100% enantioselective with l- and d-CST resulting in correspondingly left- and right-twisted bowties (Fig. 1a,c, and Supplementary Video 2); they are uniform in shape, size and handedness with standard deviation in *p*, *w*, *t* and *l* of 19.0, 13.6, 11.9 and 11.5%, respectively. Monodispersity of the bowtie particles indicates that they are formed in a self-limited assembly process (Supplementary Fig. 14)[5,6]. Nanostructured microparticles resembling a stack of flat nanoscale ‘pancakes’ are observed for *rac*-CST (Fig. 1b).

**Fig. 1: Hierarchical chirality from angstroms to micrometres via nanometers.**



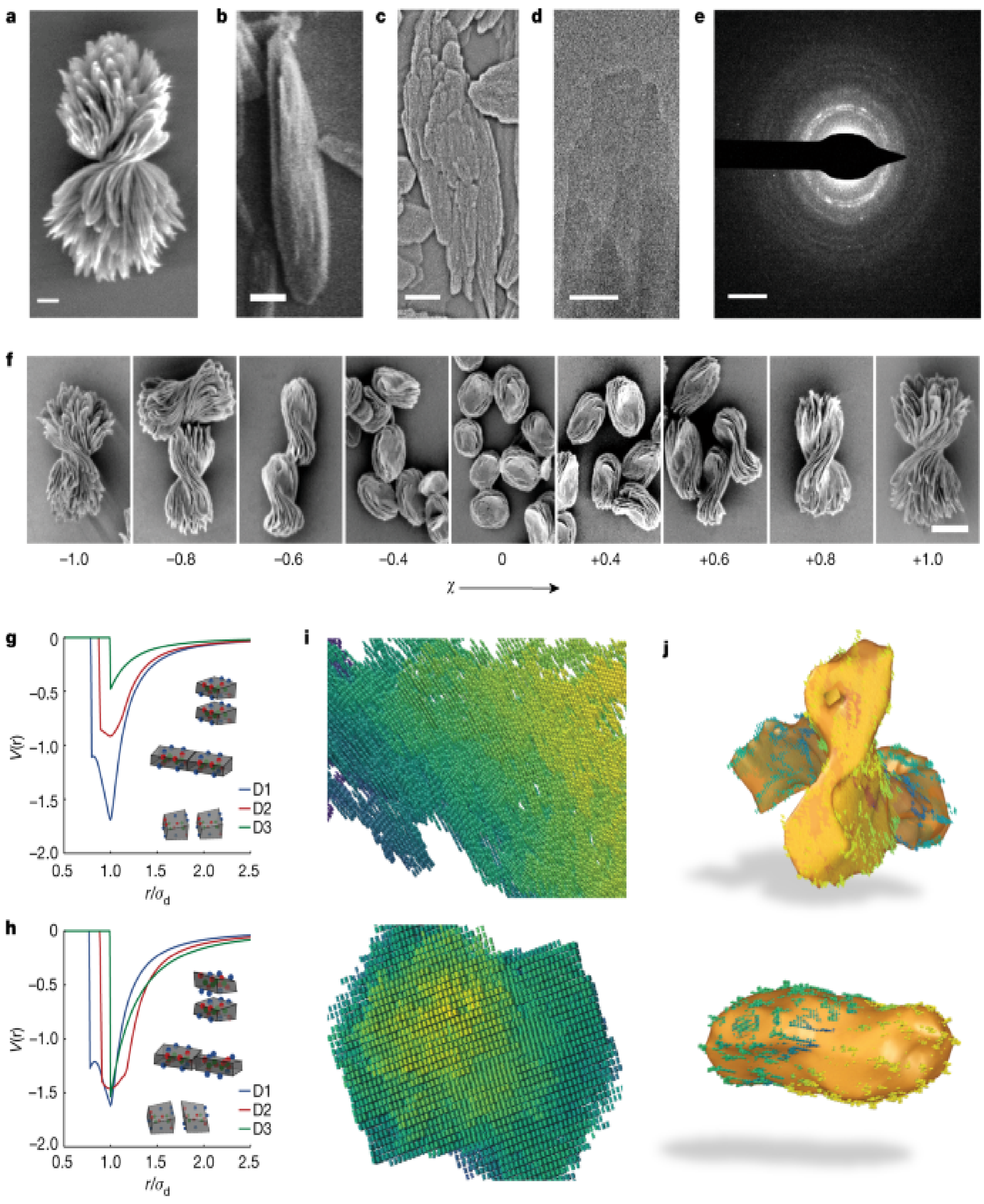
a
b
c
d
(eV)
0.0025 0.0050 0.0075
L-Bowtie
D-Bowtie
Rac-Bowtie
ΔA
ν (cm⁻¹)
A
e
(eV)
0.174 0.186 0.198 0.210
ΔA/A
ν (cm⁻¹)
A
f
(eV)
3.10 2.10 1.55 1.24
CD (mdeg)
λ (nm)
Ext_norm
g
Zeta potential (mV)
[L-CST]/[Cd²⁺]
h
Zeta potential (mV)
pH
i
× 10⁴
Intensity (AU)
Cd₂(D-CST)₂
Cd₂(L-CST)₂
Cd₂(DL-CST)₂
2θ
j
Cadmium
Nitrogen
Oxygen
Carbon
Sulfur
6.5 Å
19.6 Å
k
39.2 Å

**a**–**c**, SEM images of bowties formed from $Cd^{2+}$ and l-CST (left) (**a**), *rac*-CST (middle) (**b**) and d-CST (right) (**c**). Scale bars, 2 µm. **d**–**f**, CD (top) and extinction spectra (bottom) in THz (**d**), mid-IR (VCD) (**e**) and UV-visible-NIR (**f**) ranges for bowtie particles. **g**, Zeta potential of fully formed bowties at different ratios of l-CST and $Cd^{2+}$ ions. **h**, Zeta potential of bowties formed from l-CST solution at different pHs. **i**–**k**, Synchrotron XRD pattern (**i**) and the calculated crystal structure of nanosheets (**j**) and nanoplatelets (**k**) forming the bowties. AU, arbitrary units. The unit cell is shown in Supplementary Fig. 6.

Source data

Electron microscopy, X-ray diffraction (XRD) and electron diffraction show levels of hierarchical organization within the cadmium cystinate bowties (Figs. 1i and 2a). Scanning electron microscopy (SEM) images for the terminal (Fig. 1a–c) and intermediate (Fig. 2a,b) stages of the synthesis show that bowties are structured as a stack of twisted nanoribbons 200–1,200 nm in length and 45 nm in thickness (Supplementary Video 1). The nanoribbons are assembled from nanoplatelets 50–200 nm in length (Fig. 2c,d) and a thickness of roughly 1.2 nm calculated from the solved crystal structure (Supplementary Table 1, Supplementary Fig. 6 and Fig. 1k). Structural data in Supplementary Fig. 12 indicate that the nanoplatelets are formed from nanoclusters 1–2 nm in thickness, which is consistent with the thickness of the nanoplatelets and previous studies[28]. The nanoribbons and bowties acquire progressively stronger twists as the enantiomeric excess ($\pm\chi$) of l- or d-CST increases (Fig. 2f).

**Fig. 2: Non-binary chirality.**




**a**,**b**, SEM image of a fully formed bowtie (**a**), composed of twisted segments that are assembled from nanoribbons (**b**). **c**, SEM image of nanoribbons composed of polydisperse nanoplatelets. **d**, TEM image of the platelet. **e**. cryo-SAED pattern of the nanosheet in **d**. **f**, SEM images of the bowtie particles obtained by mixing different ratios of l-CST and d-CST, as defined

by the enantiomeric excess ($\chi$) show the transition from left-handed to pancake to right-handed bowties. Potentials of mean force obtained for preferred directions of interaction between two basic clusters of two permutations of chirality in CST ligands. **g**,**h**, LCL–LCL (**g**) and LCD–LCD (**h**) are computed and used for Monte Carlo growth simulations. The coarse-grained particle shapes deduced from the atomic structure of the nanocluster with the hydrogen bonding interaction sites marked by the spheres in LCL and LCD type nanoclusters are shown. **i**,**j**, Magnified snapshots from Monte Carlo simulations of the fully formed petal shows mismatched domains with local crystalline order arranged to form a twisted nanoribbon (**j**, top) and pancake type (**j**, bottom) petal for LCL (**i**) and LCD (**j**) type nanoclusters, respectively. Colour coding in **i** is only for visualization purposes. Scale bars, **a**, 1 µm; **b**, 0.5 µm; **c**, 200 nm; **d**, 100 nm; **e**, 2 $nm^{-1}$ and **f**, 2 µm.

Source data

Synchrotron XRD patterns helps in understanding how the atomic structure of bowties can accommodate their continuous changes in chirality while remaining chemically stable (Fig. 1a–c). Solving for the atomic structure followed by Rietveld refinement suggested several Cd-CST polymorphs (Supplementary Fig. 5–8 and Supplementary Table 1). Cumulative analysis of the XRD models points to $Cd_2CST_2$ nanoclusters (Fig. 1j) as the structural units of nanoribbons (Supplementary Fig. 8). These units produce helical chains (Fig. 1j,k and Supplementary Fig. 8) enabling partial swapping of l-CST for d-CST without drastic disruption of the short-range coordination bonds. The amine and carboxylic groups located on the exterior of the helical chains form hydrogen bonds with neighbouring chains stabilizing the structure and facilitating the nanosheet stacking (Figs. 1j,k and 2c,d). A peak overlap below 2.5 Å in the XRD data indicates imperfections in atomic organization of nanosheets that are necessary for the chemical structure to produce a multiplicity of twisted shapes.

The reasons for variable chirality are encoded in the nanoscale building blocks. First, flexible hydrogen bonds allow for variable bond angles[17,18]. Second, the ability of nanoribbons to ionize (Fig. 1g,h) leads to the long-range repulsive interactions between nanoscale building blocks that can be tuned over a wide range by changing the pH and ionic strength[29]. Because

the nanoribbons twist, the electrostatic interactions are chiral, which reinforces the handedness of their assemblies. Geometric frustrations and related mechanical deformations are contributing factors, too. In case of nanoribbons, they leed to preferential geometries at atomic, nanometer, and submicron scales[17,28,30,31], which can be captured by microscale chirality vectors, Ξ (see below).

Selected-area electron diffraction (SAED) from a single nanoribbon confirms the XRD structural analysis (Fig. 2e). Cryo-SAED patterns are consistent with corresponding distances between planes in the orthorhombic $Cd_2CST_2$ structure (Supplementary Fig. 11). The nanoribbons are polycrystalline with ordered domains of helical chains appearing as sharp spots in the SAED pattern (Fig. 2e and Supplementary Fig. 17). Some degree of polycrystallinity enables the nanoribbons to accommodate CST ligands of opposite chirality while retaining the physical integrity of the nanoassembly. 'Imperfections' and variability of nanoribbon lengths is essential for the formation of such complex particles with continuously variable twisted shape.

# Multispectral photonic activity

Extinction and CD spectroscopies from ultraviolet to terahertz (THz) frequencies demonstrate a rich set of mirror symmetric bands for the left- and right-handed bowties, whereas the 'pancake stack' assemblies are chiroptically silent (Fig. 1d,f). Visible light, Raman, Fourier transform infrared, mid-infrared and THz spectra (Fig. 1d–f) confirm the molecular structure of the bowties established by XRD and SAED. Peaks at 510 $cm^{-1}$ in Raman spectra correspond to S–S disulfide stretches (Supplementary Fig. 2a). The strong and wide band from 3,000 to 3,750 $cm^{-1}$ in Fourier transform infrared spectra confirms the multiplicity of hydrogen bonds stabilizing nanoribbons and their stacks. The CD spectrum has several positive and negative peaks denoted as P1 (−, 1,040 nm), P2 (+, 460 nm), P3 (−, 330 nm), P4 (+, 270 nm) and P5 (−, 235 nm) (Fig. 1f). P5 and P4 in the ultraviolet (UV) part of the spectrum are attributed to electronic transitions in the CST ligands. These two peaks are broadened and red-shifted compared to free CST due to coordination with $Cd^{2+}$. The vibrational CD (VCD) spectra of bowties and free CST are very different between 1,550 and

1,650 cm$^{-1}$ (Fig. 1e). The former is 100 times more intense and contains many split bands. Density functional theory calculations (Supplementary Figs. 9 and 10) identified these bands as two carbonyl stretching modes (out-of-phase and in-phase modes) coupled with H-N-H scissor oscillations. These delocalized modes with long-range couplings between the helical chains in the nanoribbons result in large absorption cross-section and drastic increase in the intensity of 1,610–1,620 cm$^{-1}$ band. The non-equivalency of the O atoms bonded to a Cd atom removes the degeneracy of the out-of-phase and in-phase modes resulting in band splitting (Figs. 1j,k and Supplementary Fig. 6). THz CD shows the existence of chiral phonons propagating along the nanoribbons (Fig. 1d)[32,33]. These data also indicate that the bowties made with l- or d-CST are true multiscale enantiomers of each other for which the mirror asymmetry is present at the subnanometre, nanometre and micrometre scales (the multiscale chirality vector below).

# Computational models

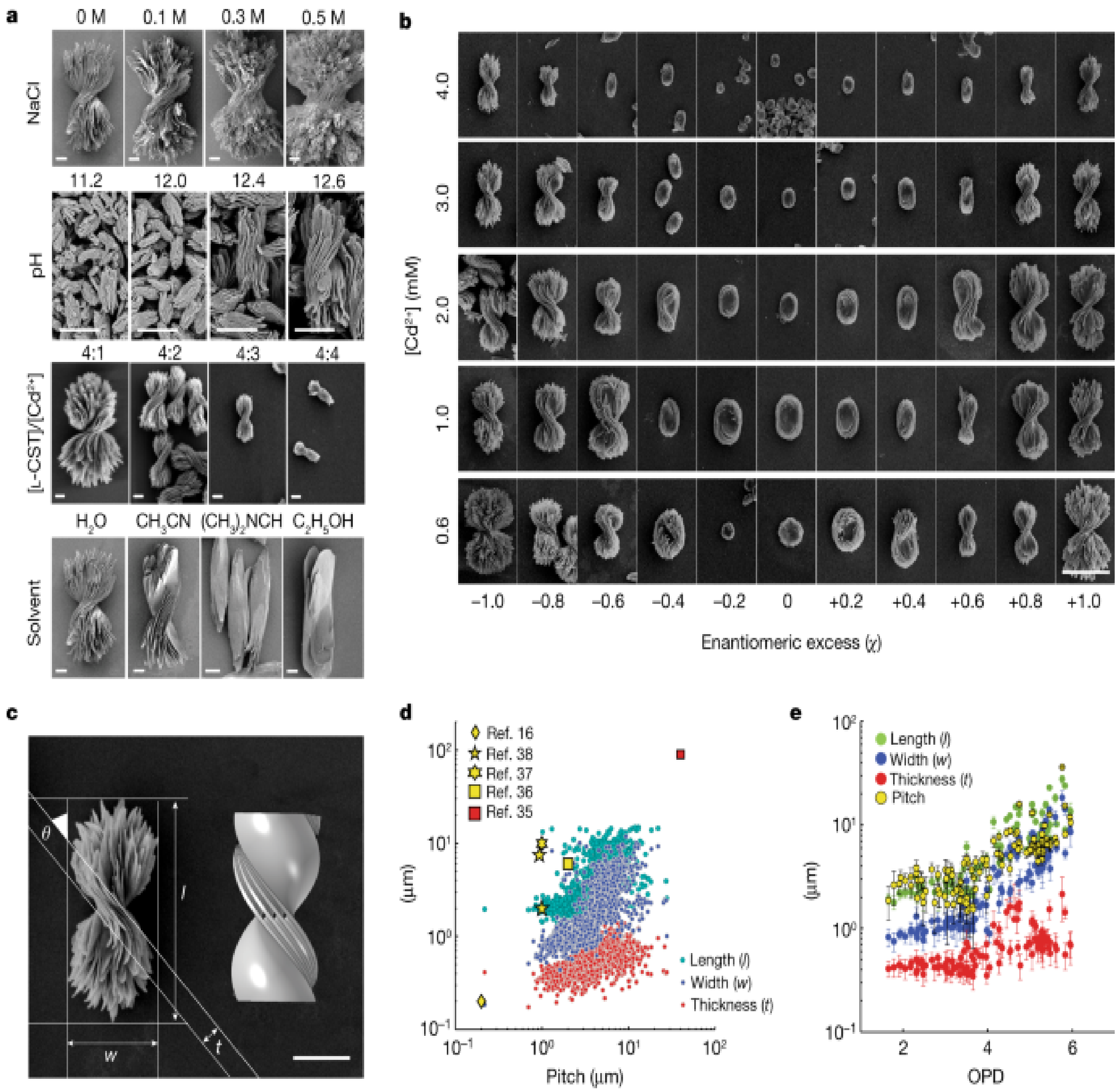
On the basis of the solved crystal structure and zeta-potential measurements (Fig. 1g,h and Supplementary Figs. 20–22), we constructed a coarse-grained version of the basic nanocluster (Supplementary Fig. 13) (Methods). A Monte Carlo growth algorithm was used to simulate bowtie assembly from nanoclusters with an average net charge of +1. Because the addition of each new nanocluster increases the total charge of the growing particle, the self-assembly process is electrostatically restricted[32], with each extra nanocluster having to overcome an increasingly repulsive energy barrier. Computation of the pair-wise potentials (Fig. 2g,h) between l-homochiral and racemic versions of the nanoclusters, abbreviated to l-CSTs (LCL) and LCD, revealed growth directions along which the nanoclusters anisotropically interact through hydrogen bonds that reflect the structural hierarchy. Nanoclusters assemble into nanoplatelets that grow into nanoribbons that stack to form the bowties. As the number of building blocks in the assembly increases, the size of the ordered domains in the nanoribbons also increases (shown in colour in Fig. 2i and Supplementary Fig. 15). The growth of nanoribbons stops when the net charge repulsion on the particle becomes sufficient to prevent further attachment of nanoclusters (Supplementary Fig. 36), balancing out the attractive forces. LCL nanoclusters form a twisted platelet (Fig. 2j, top row), whereas LCD nanoclusters form a flat sheet (Fig.

2j, bottom row). The model combining both electrostatic and geometric reconstructions correctly predicts the relationship between the enantiomeric composition of the nanoclusters, $\chi$, and the chirality of the bowties propagating up in scale as the particles evolve in size. The long-range electrostatic restrictions combined with efficacy of multiscale chirality transfer for different $\chi$ results in the uniform pitch and size of the bowties (Fig. 2f and Supplementary Fig. 31). The evidence of the elastic stress in the stacks contributing to the hierarchical twist can be appreciated by the partial unwinding of the nanoribbons on disassembly of the bowties (Supplementary Fig. 18).

# Tuneable chirality

Hierarchical assembly from chiral nanoclusters enables wide-range tuning of long- and short-range interactions by changing various physical–chemical parameters (Supplementary Fig. 25), which can be exemplified by three synthetic cases. In case 1, the increase of ionic strength results in the increase of *l*, *w* and *t* because the range of electrostatic interactions is reduced for high ionic strengths (Supplementary Fig. 20). In case 2, the zeta potential increase with pH results in the increase of *l*, *w* and *t*. De-protonation of the –COOH groups on the surface of nanoribbons and nanoplatelets results in the reduction of hydrogen bonding between the ribbons and, thus, in greater separation between the nanoribbons (Supplementary Figs. 21 and 22). In case 3, an increase of l-CST/Cd leads to an increase of *l*, *w* and *t* due to adsorption of the counter ions on the bowtie surface (Supplementary Fig. 16a). The increase of the CST concentration promotes (Fig. 3a and Supplementary Figs. 20 and 26) the close-range hydrogen bonding between the nanoclusters, resulting in thicker and wider nanoribbons (Fig. 3a and Supplementary Fig. 24).

**Fig. 3: Morphological diversity of bowties.**




**a**, Effect of NaCl, pH, excess charge (l-CST/Cd), solvent (1:1 water to solvent) on bowtie morphology (water, $[Cd^{2+}] = 1$ mM, 4 mM for pH, [l-CST] = 4 mM) are shown. Scale bar, 1 μm.**b**, Continuously variable twist and sizes for bowtie particles obtained for different $\chi$ and $[Cd^{2+}]$. Scale bar, 5 μm. **c**, SEM image of a typical bowtie assembly and the corresponding morphological parameters of length (*l*), width (*w*), thickness (*t*) and pitch ($2w\tan(90-\theta)$) overlaid. Three-dimensional models (right) used in computations of OPD chirality measures. Scale bar, 2 μm. **d**, Scatter chart demonstrating the variability of the geometrical parameters (*l, w, t*) versus pitch (*p*) for the bowtie assemblies obtained in this study and the same parameters obtained in other studies[16,36,37,38,39]; bowties retain their shape

over two orders of length scales. **e**, Variation of the morphological parameters with calculated OPD chirality measure show empirical correlations between the *l*, *w*, *t*, *p* and OPD.

Source data

The variability of the pitch serves as a demonstration of the tuneable chirality of these particles (Supplementary Fig. 33). The computational model describing the hierarchical self-assembly of the bowties from nanoclusters with intermediate formation of nanoplatelets and nanoribbons quantitatively predicts their geometry (Supplementary Figs. 31, 24 and 35). It also captures the variations of the geometrical parameters of the bowties in response to variations of ionic screening, solvents, enantiomeric excess, temperature and counter ions (Fig. 3d,e and Supplementary Figs. 25–33 and 39–43).

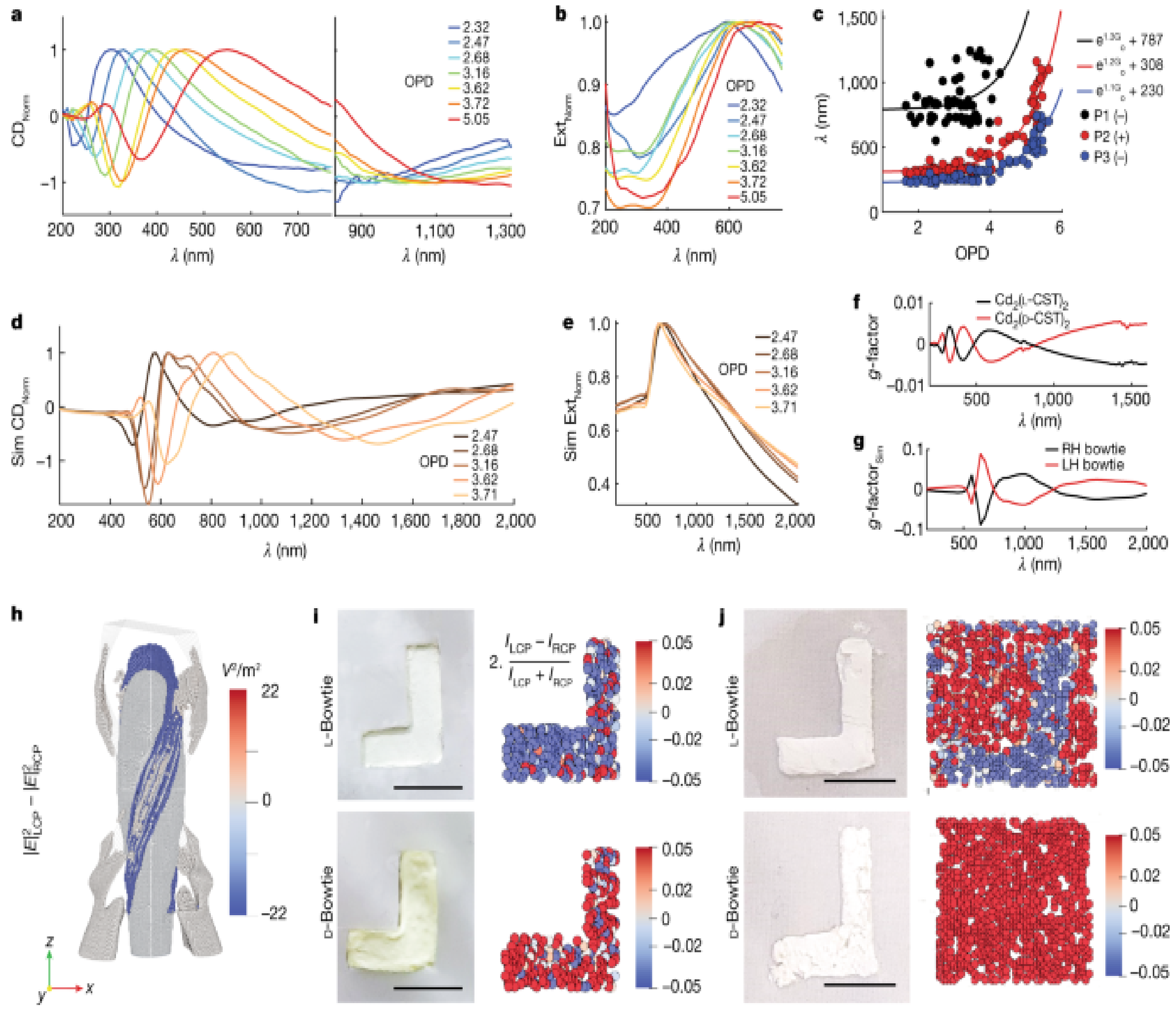
The chirality continuum can be demonstrated by calculating chirality measures, such as Hausdorff distance[1], continuous symmetry measure[2] or Osipov–Pickup–Dunmur (OPD) index (Supplementary Information)[3]. All of them are suitable for establishing the variable chirality of the bowties, but OPD has an additional useful feature: it changes sign when switching from left to right enantiomers. The transition from discrete to continuous chirality is evident from Fig. 3d,e because the pitch and OPD cover a wide range of values. For example, standard OPD varies gradually from −4.7 to +4.6, whereas other molecular or nanoscale systems have only discrete points (Fig. 3d).

Nanoassemblies have chirality at many levels, which is essential for establishing the properties of chemical structures displaying chirality continuum. Multiscale chirality can be expressed as a chirality vector $\Xi(\xi_1,\xi_2,\xi_3,\xi_4...)$ where components $\zeta_i$ represent chirality measures at specific scales. For example, $\Xi_{bowties}$ will comprise OPD for CST ligands, constituent nanoclusters, individual nanosheets (Fig. 1j,k) and bowties overall (Fig. 2a), resulting in $\Xi(\xi_i) = \Xi(-1.1 \times 10^{-3}; 1.4 \times 10^{-2}; 6.81 \times 10^{-4}; 2.75 \times 10^{-2})$; as all of them have different numbers of points, the point-scaled chirality measure $OPD_s$ instead of standard OPD was used. For

some components, i.e. $\xi_1$, chirality continuum is not observed, but for the level of fully assembled particles it is and is being considered henceforth.

Testing the correlation between the sign of peaks in the CD spectra (Figs.3e,4a and Supplementary Fig. 47d,e) is fundamentally important because of the long-standing controversy about the suitability of sign-changing pseudoscalar measures of chirality for the description of experimental data[5,34]. We found that the sign of OPD correlates with the experimental CD spectra, *g*-factor spectra and P3, P2 and P1 peaks (Fig. 4a and Supplementary Figs. 49 and 50)[5,34]. An exponential relationship between OPD and the spectral position of the P2 and P3 peaks was observed (Fig. 4c). We note that organic and biological chemistries do not provide, hitherto, examples of quantifiable property–chirality relationships, despite extensive discussions of their possibility[4,5,25,35].

**Fig. 4: Optical properties of the bowtie particles.**




**a**, Normalized CD spectra for right-handed bowties of different OPD with peaks P1(−), P2(+), P3 (−) labelled. **b**, The corresponding extinction spectra. **c**, Dependences of spectral position of P1, P2, P3 with OPD made with varying [l-CST]/[$Cd^{2+}$]. For peak P1, the dependence of its position on chirality measures is difficult to assess due to its breadth. **d**,**e**, Calculated *g* factor (**d**) and extinction (**e**) spectra are shown here for the right-handed twisted bowtie models with OPD values as in **a** and **b**. **f**, The *g*-factor spectra obtained for bowties prepared in an water-acetonitrile mixture with a peak at a 1,550 nm wavelength. **g**, Simulated *g*-factor spectra averaged for a bowtie model with *l*, *w*, *t* and *θ* as 5.3 ± 0.3 µm, 2.1 ± 0.1 µm, 0.7 ± 0.1 µm, 46 ± 5° along 25 different orientations with respect to an incident electromagnetic wave. **h**, Spatial distribution of differential ($|E|^{2}_{\rm{LCP}}-|E|^{2}_{\rm{RCP}}$) field around a bowtie model described in **g** shows enhancement of field within the layers. **i**,**j**, Printed

coatings of 1.5 × 1.5 $cm^2$ l- and d-bowties on glass (**i**) and cloth (**j**) with the point-cloud of scattered signal on illumination with 1,550 nm wavelength chiral-LIDAR laser. Scale bar, 1 cm.

Source data

# Bowtie-based photonic materials

Experimental dependences of P2/P3 peaks versus OPD indicate that the physical laws governing chirality–property relationships for nanostructured microparticles and small molecules are different. We found that the computational models describing bowties as a stack of twisted ribbons accurately explain the signs and red-shifts of P3, P2 and P1 peaks across the entire parameter space (Fig. 4a–e and Supplementary Figs. 25, 27, 28, 39–43 and 45). Decomposition of the extinction spectra from right-handed bowties into scattering and absorptive components showed that the negative P3 and P1 peaks and positive P2 peak in Fig. 1f and Fig. 4a are from differential scattering (Supplementary Fig. 49). The computational predictions for peak position match experiments nearly perfectly (Fig. 3a,b and Supplementary Figs. 26, 27, 45 and 48–50). Electromagnetic field distributions indicate that P1 and P3 originate from a mix of dipolar and quadrupolar resonance scattering modes (Fig. 4a–h and Supplementary Fig. 54). Although the peak positions vary with particle size, the attribution of the peaks remains the same for all $l$, $w$, $t$ and $\theta$, as demonstrated by the calculations for the different models (Supplementary Figs. 49 and 50). Note that both scattering and absorption peaks of bowties involve the interaction of photons with the particle as a whole, not specific quantum states, which is essential for understanding the origin of observed chirality–property relationships (Fig. 4a–e). Optical properties of molecules originate from chiral shapes of molecular orbitals that may or may not correlate with the geometry of the $sp^3$ carbon atom or Δ/Λ enantiomers. Therefore, the similarity of the physical dimensions of chiral objects in bowties and incident photons engenders the observed dependences (Fig. 4c) and, thus, the fundamental difference in chirality–property relationships in comparison with small molecules, polymers and so on.

The simplicity of bowtie synthesis and the predictive relationship between particle structure and optical properties opens the door to the engineering of particle-based metasurfaces. We demonstrate that bowties can be used as templates for inducing chiroptical emission from achiral perovskite nanocrystals (Supplementary Fig. 51). Coatings with circular polarization signatures (Supplementary Fig. 53) are needed for machine vision systems to increase the accuracy of object identification (for example, pedestrians). Dispersions of bowtie particles with strong circularly polarized P1 peak scattering in the near infrared (NIR) part of the spectrum were printed on glass slides and cotton fabric. A LIDAR operating at 1,550 nm with a circularly polarized beam (Supplementary Fig. 52) was used to acquire images of the printed objects following their decomposition into left- and right-handed components of back-scattered light originating from the P1 peak in CD spectra (Fig. 1f). The distinct contrast between prints made from the left- and right-handed bowties was observed (Fig. 4g and Supplementary Figs. 54–56). The layer of left-handed bowties can also be identified with high contrast on cotton cloth in LIDAR images (Supplementary Fig. 53d,e), which can be used in autonomous vehicles and other robots for the identification of specific materials and surface features exemplified by the polarization contrast from the micrometre-scale chirality of the fabrics and bowties.

Dispersions of bowties translate into an emerging opportunity for scalable manufacturing of printable metasurfaces. The observed continuous correlations between the position of CD peaks and OPD indicate chirality–property correlations with pseudoscalar chirality measures arise when simple chiral shapes continuously transition from left- to right-handed enantiomers through an achiral state of flat sheets by gradual twisting. Conversion of Cd-CST bowties into similarly shaped CdS particles and coating of bowties with a perovskite nanolayer (Supplementary Fig. 51) indicates that several other materials can be prepared with the same diversity of morphology (Supplementary Fig. 12). Besides machine vision, dipolar and quadrupolar photonic resonances enable chirality-based engineering of composite materials for chemical and biomedical technologies.

# Methods

## Synthesis of bowties

Cadmium chloride ($CdCl_2$), 99.99% trace metals basis, l-CST, ≥98% (TLC), crystalline, d-CST, 98%, sodium hydroxide (NaOH) pellets >97% were purchased from Sigma-Aldrich. De-ionized water (18.2 mΩ $cm^{-1}$) was used for the preparation of stock solutions and aqueous dispersions. The 10 ml stock solutions of $CdCl_2$ (0.1 M), l-CST (0.1 M), d-CST (0.1 M) and NaOH (2.5 M) were prepared by dissolving the required amounts in de-ionized water. pH of CST stock solution was raised to 11 by adding 1 ml of 2.5 NaOH in 10 ml of solution. An immediate change in appearance was observed by the formation of a clear solution. Remaining solids were dissolved by mild sonication for 10 s. Bowties were synthesized by the mixing stock solutions of $CdCl_2$ and l-CST or d-CST in a 1:1 stoichiometric ratio in an aqueous media. Typical synthesis involved 20 µl of l-CST added to 960 µl of water followed by the addition of 20 µl of $CdCl_2$. The solution mixture was shaken vigorously until the dispersion turned milky, which is an indicator of the bowties' assembly process. After that, the mixture was kept still at room temperature for 15 min to ensure the completion of the self-assembly. The dispersion was subsequently centrifuged three times in de-ionized water at 6,000 rpm for 3 min. Final aqueous dispersion was stored at room temperature and used for further characterization and studies. Bowtie dispersions for coatings were prepared by mixing the freeze-dried powder in polyacrylic acid.

## Electron microscopy

SEM samples were prepared by drop-casting 5 µl of aqueous dispersion on 1 × 1 cm silicon wafers (TedPella), following by drying at room temperature. This was followed by sputter coating a 5–10 nm film of gold on the wafer to avoid charging due to the electron beam. SEM measurements were performed in FEI Nova NanoLab Dual Beam SEM and FEI Helios Nanolab at 5 kV accelerating voltage and 0.4 nA beam current under secondary electron detection mode.

TEM samples were prepared by drop-casting 10 µl of aqueous solution on to a copper grid coated with holey carbon supported by a continuous carbon

film (TedPella 01824). Bright-field TEM was performed on JEOL 2010 operating at 300 kV accelerating voltage and acquisition were done using Gatan OneView camera. Annular dark-field–scanning transmission electron microscopy (ADF–STEM) and STEM–energy dispersive X-ray spectroscopy measurements were performed on cold-FEG JEOL 3100R05 with Cs aberration correction operating at 300 keV. A HAADF detector was used to acquire *Z*-contrast images in which the intensity is proportional to the atomic number of the column over which the electron probe is placed. Diffraction experiments were acquired using Thermo Fisher Talos F200X operated at 200 keV equipped with a Gatan OneView camera. The Elsa (698) Gatan Cryo Holder cooled specimens down to roughly 93 K for low-temperature measurements.

## STEM tomography

The synthesized bowtie particles were dispersed in water and drop cast using a micropipet onto a 3 mm copper TEM grid dried at room temperature. The TEM grid was an ultrathin (3 nm) carbon film with a large hexagonal mesh (100) to provide high specimen tilts without beam shadowing (Electron Microscopy Sciences). The tomographic tilt series were acquired using a Thermo Fisher Talos F200X operated at 200 keV as a scanning transmission electron microscope with a 10.5 mrad semiconvergence angle using a 36 and 165 mrad inner and outer semicollection angles for the ADF detector. The right-handed bowtie particles were acquired over a tilt range of −72º to +73º and left-handed bowtie particles specimen were acquired over a tilt range of −75º to +71º both with a +1º tilt increment. At each angle, ADF images of size 1,024 × 1,024 pixels were recorded with a dwell time of 4 µs and pixel size of 4.94 nm. The tomograms were reconstructed with the additive simultaneous iterative reconstruction technique for 150 iterations. The three-dimensional reconstructions were visualized by tomviz[40].

## XRD

Synchrotron X-ray powder diffraction data ($\lambda = 0.45192$ Å) were collected at Beamline 17-BM at the Advanced Photon Source at the Argonne National laboratory. A freeze-dried powder sample of l-/d-/*rac*-bowties was loaded in a 1.0 mm diameter Kapton capillary tube, which was then measured using a

VAREX XRD 4343CT amorphous silicon area detector in the Debye–Scherrer geometry.

## Measurements of particle geometry

Length ($l$), width ($w$), thickness ($h$) and angle of twist ($\theta$) as defined in Fig. 3c were measured using Jann5s or the measure tool from MATLAB central. A minimum of ten measurements were taken for each parameter and standard deviation ($\sigma$) was calculated and represented as percentage variation by following the equation ${\rm{e}}{\rm{r}}{\rm{r}}{\rm{o}}{\rm{r}}=\frac{{\sigma }_{X}}{\mathop{x}\limits_{_}}\times 100$, where $x = l$, $w$, $h$, $\theta$ and $\underline{x}$ is the mean of corresponding values.

## Finite-difference time-domain simulations of optical properties

The CD and *g*-factor spectra were calculated using a software package (Lumerical Solutions Inc.; www.lumerical.com/tcad-products/fdtd/). We used total-field scattered-field sources that surrounded the structure being modelled. Circularly polarized light was generated by positioning two total-field scattered-field sources along the same forward axis at a 90° angle and with a phase difference of either −90° for photons with left-handed polarization or 90° for photons with right-handed polarization. Two analysis groups using box power monitors monitored the absorption and scattering cross sections (extinction is the sum of absorption and scattering). The FDTD simulation region was defined by a larger box monitor with a stretched-coordinate perfectly matched layer and non-uniform mesh type. Frequency profile monitors were inserted in the total-field region to calculate electric field distribution in 3D. The refractive index for water is 1.33. Convergence tests with different mesh sizes were performed to determine the best balance between computational time restraints and simulation accuracy. Simulations were carried out for the bowties oriented with their long axis being parallel with the propagation direction (**k** vector) of photons. The 10 and 1 nm mesh sizes produced similar CD spectra; therefore, we used 10-nm mesh size for bowtie simulations presented in Supplementary Figs. 49, 50 and 54.

## Chiroptical spectroscopy

CD measurements in UV-visible and NIR range were performed using JASCO J-1700 equipped with one photomultiplier tube detector in 200–800 nm range and two InGaAs NIR detectors in 800–1,600 nm range. Typical scanning parameters were as follows: scanning speed, 500 nm $min^{-1}$; data pitch, 0.1 nm; bandwidth, 5 nm (NIR bandwidth, 10 nm), digital integration time, 0.25 s and one accumulation. The anisotropy *g* factor was calculated according to the equation $g=\frac{{\rm{C}}{\rm{D}}}{32,980\times \,{\rm{a}}{\rm{b}}{\rm{s}}}$, where CD is the signal obtained from CD/DC channel and abs is the absorbance calculated from ABS and DC channels of the spectrophotometer. CD spectra were stopped at wavelengths in which the high-tension voltage was greater than 800 V to avoid artefacts during acquisition.

## Terahertz CD spectroscopy

The terahertz (THz) time-domain polarimetry system based on three linear polarizers was used to measure THz absorption and THz CD spectra[7]. We used calculations of Stokes parameters based on the $E_x$ and $E_y$, the electric field in the $x$ and $y$ directions, respectively. THz absorptions were extracted from the equations used in ref. [7] after retrieval of phase from the transmission data. To eliminate linear birefringence effects from samples, highly concentrated mixture of bowtie powders with mineral oil was used. The reference sample is a quartz sandwich cell filled with mineral oil[32].

## VCD spectroscopy

VCD measurements were performed on freeze-dried bowtie samples dispersed in heavy water ($D_2O$) at 33 mg $ml^{-1}$ concentration. A 100 µl drop was sandwiched between two $BaF_2$ crystals separated by a 50 µm Teflon spacer. An MCT-V detector was used to acquire infrared and VCD data in the range of 2,000 to 850 $cm^{-1}$ with a resolution of 4 $cm^{-1}$ and a total of 100 and 500 accumulations, respectively. The sandwiched dispersion between $BaF_2$ crystals was rotated along the *a* axis coinciding with the direction of the beam at a constant speed to avoid settling of particles. Corresponding infrared and VCD were plotted as $A$ and $\frac{\triangle \,A}{A}$, respectively, with exclusion of the 1,300 to 1,100 $cm^{-1}$ range that

corresponds to strong absorption from $D_2O$. The simulation of the VCD spectra was performed using the method described in the Supplementary Information.

## Coarse-grained modelling of bowtie particles

We pick features consistent with XRD results to construct a coarse-grained building block for the nanoclusters. The key features we chose to adapt are (Supplementary Fig. 13) as follows. (1) Flat, rectangular prism-like building blocks with rough dimensions of 20 × 10 × 5 commensurate with XRD nanoclusters. (2) H-bonding sites decorate the two largest and two smallest facets to reflect the bonding between the –COOH and $–NH_2$ groups. (3) The two intermediate facets show a directional interaction to capture the disulfide bridge formation intrinsic to CST. These interactions sites only interact with themselves. (4) Building blocks show a slight shear on one facet to enforce chirality of l-CST versus R-CST. (5) Isotropic charge–charge repulsion between $Cd^{2+}$ atoms embedded inside the building block: each has a net charge of +1 to build in charge-driven, self-limiting behaviours. (6) All other surface sites not defined above show steric repulsion.

## Potential of mean force and Monte Carlo growth simulation

Potential of mean force (PMF) values between nanoclusters are computed by adding together all interactions as a function of the centre-to-centre distance between nanocluster pairs and performing weighted averaging. The resulting PMFs between LCL highlight directional growth features that are commensurate with the hierarchy of structural formation observed in experiments (Fig. 2g). PMFs computed for a mix of l-CST and r-CST (LCD) conversely suggest a reduction in hierarchical structural formation and chiral propagation (Fig. 2h). We then use the computed PMFs to perform Monte Carlo growth simulations. Specifically, the well depth of each computed PMF plus the net charge–charge repulsion between the next nanocluster and growing structure are used as the input energy in the Metropolis algorithm. This enables Boltzmann weighted selection of the direction in which to place the next building block in the growing structure. Visualization of the

growth simulation results confirm the formation of bowtie and pancake morphologies (Fig. 2i,j).

## Data availability

The authors declare that the data supporting the findings of this study are available within the article and its supplementary information files. Source data are provided with this paper.

## References


1. Buda, A. B. & Mislow, K. A Hausdorff chirality measure. *J. Am. Chem. Soc.* **114**, 6006–6012 (1992).

2. Osipov, M. A., Pickup, B. T. & Dunmur, D. A. A new twist to molecular chirality: intrinsic chirality indices. *Molec. Phys.* **84**, 1193–1206 (1995).

3. Zabrodsky, H. & Avnir, D. Continuous symmetry measures. 4. Chirality. *J. Am. Chem. Soc.* **117**, 462–473 (1995).

4. Rassat, A. & Fowler, P. W. Is there a 'most chiral tetrahedron'? *Chemistry* **10**, 6575–6580 (2004).

5. Harris, A. B., Kamien, R. D. & Lubensky, T. C. Molecular chirality and chiral parameters. *Rev. Mod. Phys.* **71**, 1745–1757 (1999).

6. Xia, Y. et al. Self-assembly of self-limiting monodisperse supraparticles from polydisperse nanoparticles. *Nat. Nanotechnol.* **6**, 580–587 (2011).

7. Choi, W. J. et al. Terahertz circular dichroism spectroscopy of biomaterials enabled by kirigami polarization modulators. *Nat. Mater.* **18**, 820–826 (2019).

8. Nguyen, M.-K. & Kuzyk, A. Reconfigurable chiral plasmonics beyond single chiral centers. *ACS Nano* **13**, 13615–13619 (2019).

9. Dey, S. et al. DNA origami. *Nat. Rev. Methods Prim.* **1**, 13 (2021).

10. Probst, P. T. et al. Mechano-tunable chiral metasurfaces via colloidal assembly. *Nat. Mater.* **20**, 1024–1028 (2021).

11. Kim, Y. et al. Reconfigurable chiroptical nanocomposites with chirality transfer from the macro- to the nanoscale. *Nat. Mater.* **15**, 461–468 (2016).

12. Miyajima, D. et al. Liquid crystalline corannulene responsive to electric field. *J. Am. Chem. Soc.* **131**, 44–45 (2009).

13. Ohta, E. et al. Redox-responsive molecular helices with highly condensed π-clouds. *Nat. Chem.* **3**, 68–73 (2011).

14. Ma, W. et al. Chiral plasmonics of self-assembled nanorod dimers. *Sci. Rep.* **3**, 1934 (2013).

15. Yashima, E. et al. Supramolecular helical systems: helical assemblies of small molecules, foldamers, and polymers with chiral amplification and their functions. *Chem. Rev.* **116**, 13752–13990 (2016).

16. Lee, H. E. et al. Amino-acid- and peptide-directed synthesis of chiral plasmonic gold nanoparticles. *Nature* **556**, 360–364 (2018).

17. Jiang, W. et al. Emergence of complexity in hierarchically organized chiral particles. *Science* **368**, 642–648 (2020).

18. Medina, D. D. & Mastai, Y. Chiral polymers and polymeric particles for enantioselective crystallization. *Isr. J. Chem.* **58**, 1330–1337 (2018).

19. Gingras, M. One hundred years of helicene chemistry. Part 3: applications and properties of carbohelicenes. *Chem. Soc. Rev.* **42**, 1051–1095 (2013).

20. Ma, W. et al. Chiral inorganic nanostructures. *Chem. Rev.* **117**, 8041–8093 (2017).

21. Aida, T., Meijer, E. W. & Stupp, S. I. Functional supramolecular polymers. *Science* **335**, 813–817 (2012).

22. Ho, R.-M. et al. Block copolymers with a twist. *J. Am. Chem. Soc.* **131**, 18533–18542 (2009).

23. Wang, H.-F. et al. Generalizing the effects of chirality on block copolymer assembly. *Proc. Natl Acad. Sci. USA* **116**, 4080 LP–4084089 (2019).

24. Ido, S. et al. Beyond the helix pitch: direct visualization of native DNA in aqueous solution. *ACS Nano* **7**, 1817–1822 (2013).

25. Petitjean, M. Chirality and symmetry measures: a transdisciplinary review. *Entropy* https://doi.org/10.3390/e5030271 (2003).

26. Buda, A. B., der Heyde, T. A. & Mislow, K. On quantifying chirality. *Angew. Chemie Int. Ed. English* **31**, 989–1007 (1992).

27. Kotov, N. A., Fendler, J. H. & Dekany, I. Layer-by-layer self-assembly of polyelectrolyte- semiconductor nanoparticle composite films. *J. Phys. Chem.* **99**, 13065–13069 (1995).

28. Tang, Z., Zhang, Z., Wang, Y., Glotzer, S. C. & Kotov, N. A. Self-assembly of CdTe nanocrystals into free-floating sheets. *Science* **314**, 274–278 (2006).

29. Nguyen, T. D., Schultz, B. A., Kotov, N. A. & Glotzer, S. C. Generic, phenomenological, on-the-fly renormalized repulsion model for self-limited organization of terminal supraparticle assemblies. *Proc. Natl Acad. Sci. USA* **112**, E3161–E3168 (2015).

30. Gao, C. et al. Electrostatic shape control of a charged molecular membrane from ribbon to scroll. *Proc. Natl Acad. Sci. USA* **116**, 22030–22036 (2019).

31. Miller, J. M. et al. Conformational switching of chiral colloidal rafts regulates raft–raft attractions and repulsions. *Proc. Natl Acad. Sci. USA* **116**, 15792–15801 (2019).

32. Choi, W. J. et al. Chiral phonons in microcrystals and nanofibrils of biomolecules. *Nat. Photonics.* **16**, 366-373 (2022).

33. Yeom, J. et al. Chiromagnetic nanoparticles and gels. *Science* **359**, 309–314 (2018).

34. Millar, G., Weinberg, N. & Mislow, K. On the Osipov–Pickup–Dunmur chirality index: why pseudoscalar functions are generally unsuitable to quantify chirality. *Mol. Phys.* **103**, 2769–2772 (2005).

35. Pelayo, J. J., Whetten, R. L. & Garzón, I. L. Geometric quantification of chirality in ligand-protected metal clusters. *J. Phys. Chem. C.* **119**, 28666–28678 (2015).

36. Shaw, L. A. et al. Scanning two-photon continuous flow lithography for synthesis of high-resolution 3D microparticles: erratum. *Opt. Express* **26**, 14718 (2018).

37. Chang, Y.-H., Jang, J.-W., Chang, Y.-C., Lee, S.-H. & Siao, T.-F. Gold nanohelices: a new synthesis route, characterization, and plasmonic E-field enhancement. *ACS Omega* **5**, 14860–14867 (2020).

38. Kaschke, J. & Wegener, M. Gold triple-helix mid-infrared metamaterial by STED-inspired laser lithography. *Opt. Lett.* **40**, 3986 (2015).

39. Yan, J. et al. Self-assembly of chiral nanoparticles into semiconductor helices with tunable near-infrared optical activity. *Chem. Mater.* **32**, 476–488 (2020).

40. Levin, B. D. A. et al. Tutorial on the visualization of volumetric data using tomviz. *Micros. Today* **26**, 12–17 (2018).

# Acknowledgements

This project was supported by the Vannevar Bush DoD Fellowship to N.A.K. titled 'Engineered Chiral Ceramics' Office of Naval Research (ONR), grant no. N000141812876, National Science Foundation (NSF) project 'Energy- and Cost-Efficient Manufacturing Employing

Nanoparticles’, grant nos. NSF 1463474, ONR N000141812876; ONR COVID-19 Newton Award ‘Pathways to Complexity’ with ‘Imperfect’ Nanoparticles’, grant no. HQ00342010033, NSF 1566460 ‘Nanospiked Particles for Photocatalysis’; NSF grant no. 1538180 and NSF grant no. DMR-9871177 for funding the JEOL 2010F analytical electron microscope used in this work. N.A.K., P.K. M.V. and J.S. are also thankful to the Office of Naval Research for the support from a MURI project  Mechanical Metamaterials, ONR grant no. N00014-20-1-2479. Support for the Dual Source and Environmental X-ray Scattering Facility at the University of Pennsylvania was provided by the Laboratory for Structure and Matter, which is funded in part by NSF grant no. MRSEC 1720530. P.K., T.V., S.G., E.M., C.M. and N.A.K. were supported by Office of Naval Research Multidisciplinary University Research Initiative Award no. ONR N00014-18-1-2497; R.H. acknowledges support from Army Research Office (ARO), grant no. W911NF-22-1-0056. This research was supported in part through computational resources and services provided by Advanced Research Computing, a division of Information and Technology Services at the University of Michigan, Ann Arbor, USA. V.P.N. acknowledges funding from Unitatea Executiva pentru Finantarea Invatamantului Superior, a Cercetarii, Dezvoltarii si Inovarii (UEFISCDI), grant no. PN-III-P4-ID-PCE-2020-2783-P, contract no. PCE 2/2022. Initial experiments by J. Yan and Z. Mu are gratefully acknowledged[39]. We are also grateful to G. Grason, A. Hunter and B. Palen for helpful discussions.

# Author information

## Authors and Affiliations

1. Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA

   Prashant Kumar, Thi Vo, Minjeong Cha, Anastasia Visheratina, Ji-Young Kim, Alexander Simon, Daniel Katz, Won Jin Choi, Michael Veksler, Sharon Glotzer & Nicholas A. Kotov

2. Biointerfaces Institute, University of Michigan, Ann Arbor, MI, USA

Prashant Kumar, Thi Vo, Minjeong Cha, Anastasia Visheratina, Ji-Young Kim, Won Jin Choi, Sharon Glotzer & Nicholas A. Kotov

3. X-ray Science Division, Advanced Photon Source, Argonne National Laboratory, Lemont, IL, USA

Wenqian Xu & Si Chen

4. Department of Materials Science and Engineering, University of Michigan, Ann Arbor, MI, USA

Jonathan Schwartz, Robert Hovden, Sharon Glotzer & Nicholas A. Kotov

5. Pro-Vitam Ltd, Sfantu Gheorghe, Romania

Valentin Paul Nicu

6. Department of Chemistry, University of Pennsylvania, Philadelphia, PA, USA

Emanuele Marino & Christopher Murray

7. Department of Physics and Chemistry, University of Palermo, Palermo, Italy

Emanuele Marino

## Contributions

N.A.K. designed and directed all aspects of the project. P.K. conducted the synthesis of libraries of bowties with assistance from A.S. and D.K. P.K. measured VCD and CD data. W.J.C. measured the terahertz CD spectroscopy data. P.K., W.X. and S.C. acquired, solved and refined the crystal structure from XRD data with input from N.A.K. P.K. conducted all the SEM imaging. J.S., P.K. and R.H. conducted the TEM and tomography measurements. T.V. and S.G. conducted the Monte Carlo growth simulations and theoretical model development for structure prediction with input from P.K. and N.A.K. J.-Y.K. and M.C. wrote the code for calculation of OPD

indices and J.-Y.K. calculated all the values. A.V. built 3D models of bowties based on geometrical parameters observed in SEM images for OPD calculation. V.P.N. and P.K. conducted the VCD simulations. P.K. simulated the chiroptical properties with input from N.A.K. E.M. and C.M. measured single particle scattering data. M.V. developed perovskite coated bowties. M.C. and N.A.K. conducted the LIDAR measurement and analysis. All the authors discussed the results. P.K. and N.A.K. wrote the paper with written contributions from all the authors. N.A.K. put together chirality vectors Ξ.

### Corresponding authors

Correspondence to Sharon Glotzer or Nicholas A. Kotov.

## Ethics declarations

### Competing interests

N.A.K. is a founder of a start-up company Photon Semantics developing LIDAR technologies for machine vision. The other authors declare no competing interests.

## Peer review

### Peer review information

*Nature* thanks Tobias König and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.Peer reviewer reports are available.

## Additional information

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Supplementary information

**Supplementary Information**

This file contains Supplementary Methods, Figs. 1–56, Tables 1–3 and References.

**Supplementary Data**

Python code for OPD calculations.

**Peer Review File**

**Supplementary Video 1**

Slice and view of the bowtie particle.

**Supplementary Video 2**

Tomogram of the bowtie particle.

## Source data

**Source Data Fig. 1**

**Source Data Fig. 2**

**Source Data Fig. 3**

**Source Data Fig. 4**

## Rights and permissions

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted

manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Reprints and Permissions

# About this article

## Cite this article

Kumar, P., Vo, T., Cha, M. *et al.* Photonically active bowtie nanoassemblies with chirality continuum. *Nature* **615**, 418–424 (2023). https://doi.org/10.1038/s41586-023-05733-1

- Received: 01 May 2022
- Accepted: 16 January 2023
- Published: 15 March 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05733-1

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

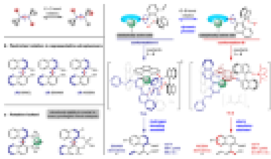
## Bow-tie particles boast a tunable twist

- Bart Kahr

# Nature News & Views 15 Mar 2023

---

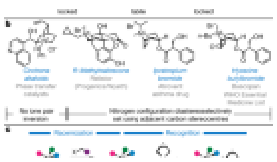
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05733-1



## Conformational enantiodiscrimination for asymmetric construction of atropisomers

12 August 2022

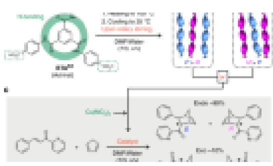
Shouyi Cen, Nini Huang, … Zhipeng Zhang



## Enantioselective synthesis of ammonium cations

01 September 2021

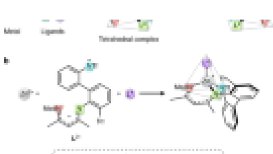
Mark P. Walsh, Joseph M. Phelps, … Matthew O. Kitching



## Asymmetric catalysis mediated by a mirror symmetry-broken helical nanoribbon

04 September 2019

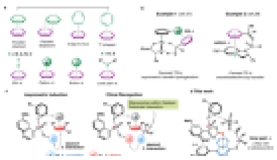
Zhaocun Shen, Yutao Sang, … Minghua Liu



[**Asymmetric construction of tetrahedral chiral zinc with high configurational stability and catalytic activity**]()

09 December 2020

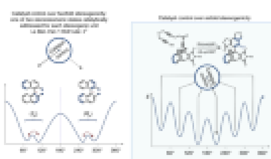
Kenichi Endo, Yuanfei Liu, … Mitsuhiko Shionoya



[**Engineered non-covalent π interactions as key elements for chiral recognition**]()

07 June 2022

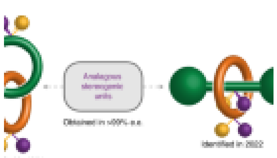
Ming Yu Jin, Qianqian Zhen, … Chen Xu



[**Catalyst control over sixfold stereogenicity**]()

27 May 2021

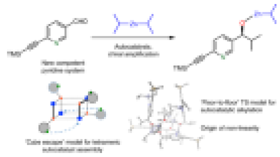
Xingxing Wu, Reto M. Witzig, … Christof Sparr



[**Mechanically axially chiral catenanes and noncanonical mechanically axially chiral rotaxanes**]()

27 June 2022

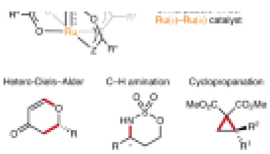
John R. J. Maynard, Peter Gallagher, … Stephen M. Goldup



## [Demystifying the asymmetry-amplifying, autocatalytic behaviour of the Soai reaction through structural, mechanistic and computational studies]

23 March 2020

Soumitra V. Athavale, Adam Simon, … Scott E. Denmark



## [Chiral paddle-wheel diruthenium complexes for asymmetric catalysis]

05 October 2020

Taku Miyazawa, Takuro Suzuki, … Shigeki Matsunaga

- Article
- Open Access
- Published: 15 March 2023

# Control of stereogenic oxygen in a helically chiral oxonium ion


- Owen Smith,
- Mihai V. Popescu,

- Madeleine J. Hindson,
- Robert S. Paton ,
- Jonathan W. Burton &
- …
- Martin D. Smith

Show authors

*Nature* **volume 615**, pages 430–435 (2023)

- 5827 Accesses
- 81 Altmetric
- Metrics details

# Abstract

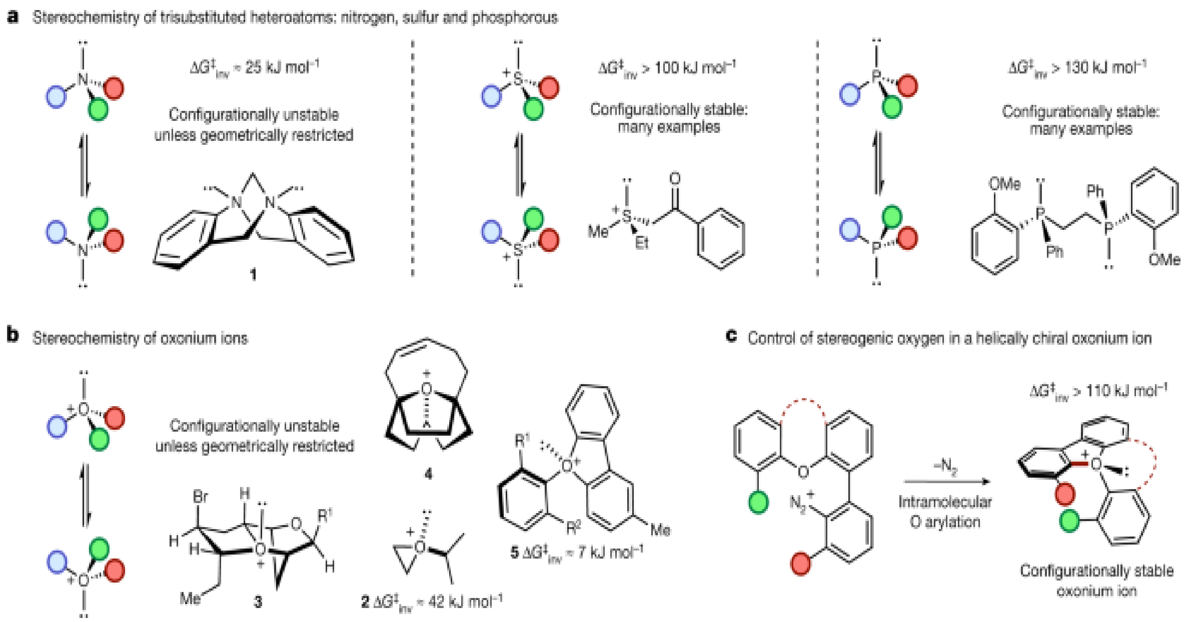
The control of tetrahedral carbon stereocentres remains a focus of modern synthetic chemistry and is enabled by their configurational stability. By contrast, trisubstituted nitrogen[1], phosphorus[2] and sulfur compounds[3] undergo pyramidal inversion, a fundamental and well-recognized stereochemical phenomenon that is widely exploited[4]. However, the stereochemistry of oxonium ions—compounds bearing three substituents on a positively charged oxygen atom—is poorly developed and there are few applications of oxonium ions in synthesis beyond their existence as reactive intermediates[5,6]. There are no examples of configurationally stable oxonium ions in which the oxygen atom is the sole stereogenic centre, probably owing to the low barrier to oxygen pyramidal inversion[7] and the perception that all oxonium ions are highly reactive. Here we describe the design, synthesis and characterization of a helically chiral triaryloxonium ion in which inversion of the oxygen lone pair is prevented through geometric restriction to enable it to function as a determinant of configuration. A combined synthesis and quantum calculation approach delineates design principles that enable configurationally stable and room-temperature isolable salts to be generated. We show that the barrier to inversion is greater than 110 kJ $mol^{-1}$ and outline processes for resolution. This constitutes, to our

knowledge, the only example of a chiral non-racemic and configurationally stable molecule in which the oxygen atom is the sole stereogenic centre.

# Main

The pyramidal inversion of trisubstituted amines, phosphines and sulfonium salts is a fundamental stereochemical process that has been intensively studied, leading to an understanding of the factors that affect the magnitude of their inversion barriers[1]. This has enabled the generation of a myriad of configurationally stable chiral phosphines (which possess barriers to inversion of 120–185 kJ mol$^{-1}$)[3] and sulfonium salts (with barriers of 100–130 kJ mol$^{-1}$)[4]. The inversion of the nitrogen atom in a tertiary amine has a relatively low barrier (20–25 kJ mol$^{-1}$) and there is a substantial contribution from quantum tunnelling to the observed rate[2]; this means that the preparation of configurationally stable tertiary amines bearing a stereogenic nitrogen atom is particularly challenging. However, there are numerous examples of bicyclic amines such as Tröger's base **1**[8,9,10], cinchona alkaloids[11] and others[12,13] in which nitrogen inversion is restricted by virtue of conformational rigidity, enabling access to compounds bearing a stereogenic nitrogen (Fig. 1a)[14]. By contrast, the study of oxonium ions is much less developed[15,16]. The only empirical measurement of the inversion barrier for a trivalent oxygen compound is for the epoxide-derived oxonium salt **2** that has a barrier to inversion of 42 kJ mol$^{-1}$, which is substantially below that required to maintain configurational stability at ambient temperature[7]. Consequently, there are no examples of the isolation of enantioenriched oxonium ions in which the oxygen atom[17,18] is the sole stereogenic centre (Fig. 1b).

**Fig. 1: Pyramidal inversion of trisubstituted heteroatoms.**




**a**, Pyramidal inversion of common trisubstituted heteroatoms occurs at different rates, leading to different configurational stabilities for tertiary amines, sulfonium salts and phosphines. **b**, Inversion of oxonium ions is rapid unless prevented by geometrical restriction: there are no examples of room-temperature configurationally stable oxonium ions bearing oxygen as the sole stereogenic centre. **c**, Synthesis and characterization of configurationally stable oxonium ions bearing a stereogenic oxygen. Counterions are not shown for clarity.$\Delta G^{\ddagger}_{inv}$ = Gibbs free energy of activation for the inversion process.

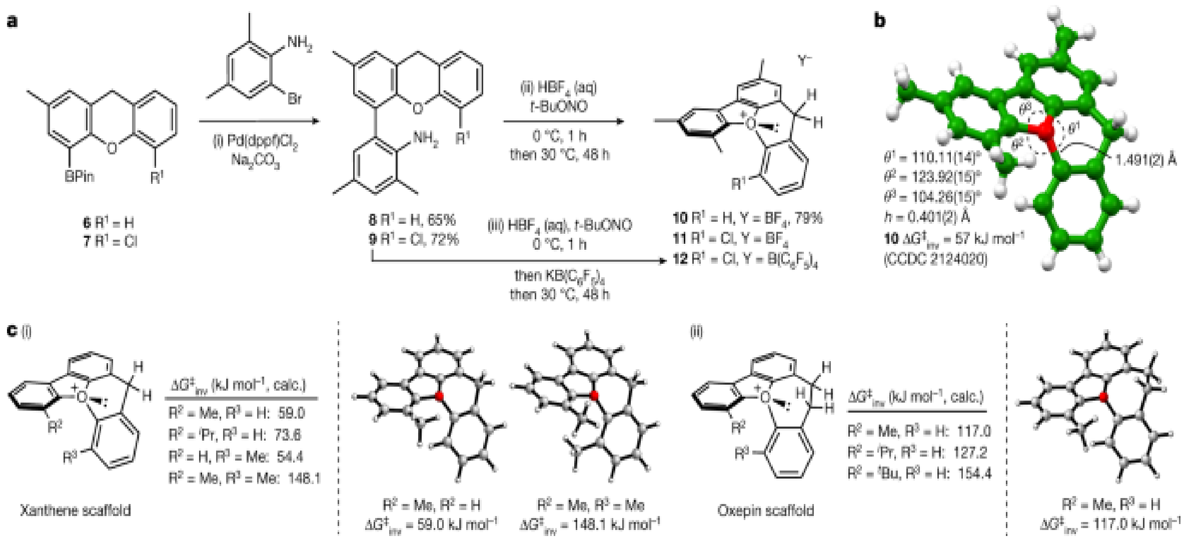
Oxonium ions are considered to be highly reactive and are invoked as transient intermediates in a range of synthetic transformations including natural product biosynthesis and total synthesis[19,20,21,22,23,24]. This has led to the characterization of complex trialkyloxonium ions such as **3**, a proposed intermediate in the synthesis of the Laurencia family of natural products. The fused ring system in **3** prevents inversion, and hence the oxygen is stereogenic; in this case the oxonium ion is formed in a diastereoselective manner from a precursor bearing several carbon stereocenters[5]. Tetracyclic oxonium ion **4** possesses a stereogenic oxygen due to its non-symmetrical bridge[25]. The best-known oxonium ions are arguably the trialkyl tetrafluoroborates, first synthesized by Meerwein more than 80 years ago[6,26]. Although these are stable and isolable, they are also some of the most reactive alkylating agents used in a synthetic chemistry laboratory and react

readily with weak nucleophiles including water. Other alkyl oxonium ions have been prepared and studied[27,28,29,30], but species stable to nucleophiles were elusive until Mascal reported oxatriquinane oxonium ions[31,32,33,34]; this indicates the potential impact upon stability of embedding oxonium ions in ring structures. In contrast to most alkyl-substituted oxonium ions, simple triaryloxonium ions are almost trigonal planar about oxygen and are stable isolable solids resistant to the action of many nucleophiles[35,36,37]. Dibenzofuran-derived triaryloxonium ions are pyramidal about the oxygen atom[38], which led to Baldridge and Siegel investigating rotation about the exocyclic C–O bond in oxonium salts such as **5** ($R^1 = Et$, $R^2 = H$); this showed that the barrier (estimated at more than 84 kJ $mol^{-1}$) was probably relatively close to Oki's definition of atropisomerism[39]. However, oxygen inversion was calculated to require substantially less energy ($R^1 = R^2 = Me \approx$ 7 kJ $mol^{-1}$)[40]. We reasoned that joining of the aromatic rings in the triaryloxonium scaffold with linkers of various lengths and flexibilities should increase the barrier to oxygen inversion while also maintaining the stability exhibited by other triaryloxonium ions (Fig. 1c). A sufficiently high enantiomerization barrier (which we have defined as greater than 107 kJ $mol^{-1}$, $t_{1/2}$ rac = 7 days at 25 °C; $t_{1/2}$ rac = the time during which the enantiomeric purity of a chiral sample has reduced from 100 to 50% enantiomeric excess) would effectively lock the configuration of the stereogenic oxygen lone pair on a laboratory timescale and generate helically chiral molecules[41,42]. We reasoned that an intramolecular O-arylation of a dibenzofuran scaffold with a diazonium salt could offer a viable route to such molecules.

We initially selected a dibenzofuran−xanthene scaffold for our investigations as the $CH_2$ protons in the xanthene backbone could be used as a reporter group to measure inversion barriers by variable temperature $^1H$ NMR spectroscopy. We prepared the functionalized xanthenes **6** and **7** from commercial materials (see page 3 of the Supplementary Information for details); these were cross-coupled with 4,6-dimethyl-2-bromoaniline to afford biaryls **8** and **9**. Diazotization and intramolecular O-arylation with a loss of nitrogen led to the formation of pentacyclic oxonium tetrafluoroborates **10** and **11** (Fig. 2a). The structure of compound **10** was confirmed by single-crystal X-ray diffraction, which shows that the

compound is pyramidal at the oxygen (sum of the angles around the oxygen $\theta^1 + \theta^2 + \theta^3 = 338.3(3)°$ versus 345(1)° for the dibenzofuran parent oxonium[43], and the apex height $h = 0.401(2)$ Å, Fig. 2) with the longest C–O bond length of 1.491(2) Å. The dibenzofuran ring is planar (internal C–O–C angle 104.26 (15)°), with the central ring of the xanthene adopting a boat-like conformation that places its two adjoining aromatic rings in different planes (Fig. 2b). The xanthene C–O–C internal angle is 110.11(14)° (versus 119.3° in the parent heterocycle xanthene itself), indicative of some angle strain at the fusion with the benzofuran ring, although that does not seem to compromise the stability of the compound, which can be stored indefinitely at room temperature without special precautions. Low-temperature $^1$H NMR spectroscopy was used to determine the enantiomerization barrier as 58.3 kJ mol$^{-1}$ (see page 23 of the Supplementary Information for further details), which is substantially lower than the estimated value of 107 kJ mol$^{-1}$ required to enable configurationally stable molecules to be isolated. By contrast, oxonium ion **11** was only observed as a transient species, and we initially reasoned that this may have been a consequence of the tetrafluoroborate counterion functioning as a nucleophilic fluoride source (the Balz−Schiemann reaction). Anion exchange of the diazonium tetrafluoroborate salt derived from **9** to a less nucleophilic $B(C_6F_5)_4^-$ counterion and ring closure afforded **12**, but this compound slowly decomposed in solution at ambient temperature over 72 hours. We considered that the introduction of a second substituent onto the dibenzofuran−xanthene scaffold may have increased the strain in **11** and **12**, compromising the relative stability observed in **10**.

**Fig. 2: Synthetic route to and quantum chemical calculations of first-generation stereogenic-at-oxygen compounds.**




**a**, Reagents and conditions: (i) $Pd(dppf)Cl_2{\cdot}CH_2Cl_2$ (5 mol%), $Na_2CO_3$ (4 equivalents), PhMe/$H_2O$ (1:1, v/v), 95 °C, 24 h; (ii) $HBF_4$ (48% aq, 5 equivalents), *t*-BuONO (5 equivalents), $CH_2Cl_2$/IPA (1:1), 0 °C, 1 h then 30 °C, 48 h; (iii) $HBF_4$ (48% aq, 5 equivalents), *t*-BuONO (5 equivalents), $CH_2Cl_2$/IPA (1:1), 0 °C, 1 h, then $KB(C_6F_5)_4$, then 30 °C, 48 h. **b**, Crystal and molecular structure of fused pentacyclic oxonium salt **10** from single-crystal X-ray diffraction studies. The tetrafluoroborate counterion is omitted for clarity. Numbers in parentheses indicate standard uncertainties. **c**, B3LYP-D3(BJ)/6.31+G(d,p) computed inversion barriers for a series of oxonium ions based upon a dibenzofuran-xanthene core (i) and a dibenzofuran-benzooxepine core (ii). *h*, Height of conceptualized triangular oblique pyramid with oxygen at its apex and the adjacent carbons of the attached O-aryl rings forming the base; Calc., calculated.

To enable us to rationalize the magnitude of the inversion barrier in **10** and the reactivity of more substituted compounds such as **11** we applied B3LYP-D3(BJ)/6-31+G(d,p) quantum chemical calculations (see page 33 of the Supplementary Information). We initially calculated barriers to the inversion of the oxonium ions based around the benzofuran−xanthene scaffold (Fig. 2c(i)). This gave excellent agreement with the experimental values for the overall geometry and barrier to pyramidal inversion for a model 2-methyl oxonium ion ($R^2$ = Me, $R^3$ = H: calculated at 59.0 kJ $mol^{-1}$); other levels of theory and basis sets gave comparable quantitative results. These calculations also indicated that increasing the size of the substituent $R^2$

would result in an increased inversion barrier (73.6 kJ mol$^{-1}$ for R$^2$ = $^i$Pr, R$^3$ = H) and a marginal increase in pyramidalization about oxygen. Switching the (methyl) substitution to the xanthene portion gave predicted enantiomerization barriers similar to that previously calculated (R$^2$ = H, R$^3$ = Me, 54.4 kJ mol$^{-1}$), but below the desired level for configurational stability. To investigate the impact of the addition of substituents on both rings, we examined a doubly alkyl-substituted oxonium (R$^2$ = R$^3$ = Me). This led to an increase in the predicted barrier to enantiomerization to 148.1 kJ mol$^{-1}$ but also led to some changes in geometry that may explain the observed instability of molecules such as **11** and **12** at ambient temperature. The calculated sum of angles around the oxonium oxygen is 345.1° (R$^2$ = R$^3$ = Me), similar to the parent dibenzofuran oxonium ion, but we observe a notable increase in the external C–O–C angle to 129° and a reduction in apex height ($h$ = 0.33 Å) consistent with a reduction in pyramidalization around oxygen. We also observe further lengthening of the longest C–O bond to 1.51 Å. This is a result of steric repulsion between the alkyl substituents, which leads to a twist in the xanthene ring and deformation from the boat-like conformation populated in **10**. Calculating a series of hyperhomodesmotic bond separation reaction enthalpies[44] enabled us to estimate the increase in steric strain in going from a compound such as **10** (R$^1$ = H) to one bearing an extra alkyl group (R$^1$ = Me) as 16.7 kJ mol$^{-1}$ (see page 34 of the Supplementary Information for details). For **11**, the calculated enthalpy difference versus **10** is significantly higher (45.6 kJ mol$^{-1}$), probably indicating both an electronic and a steric contribution. We can conclude that the development of oxonium ions with large enough inversion barriers to attain configurational stability through the addition of multiple substituents on this ring system without compromising stability will probably be a challenge.

The design of configurationally stable molecules requires that we maximize the difference between ground states (GSs) and transition states (TSs). With a six-membered ring, our attempts to destabilize the TS (through installation of a larger R$^3$ substituent) also results unavoidably in GS destabilization. Dihydrodibenzooxepines have a non-planar structure and we envisaged that the extra sp$^3$ carbon in the seven-membered ring (versus xanthene) may enable the avoidance of steric destabilization in the GS. In contrast, we reasoned that as the molecule moves towards planarity in the TS, the extra

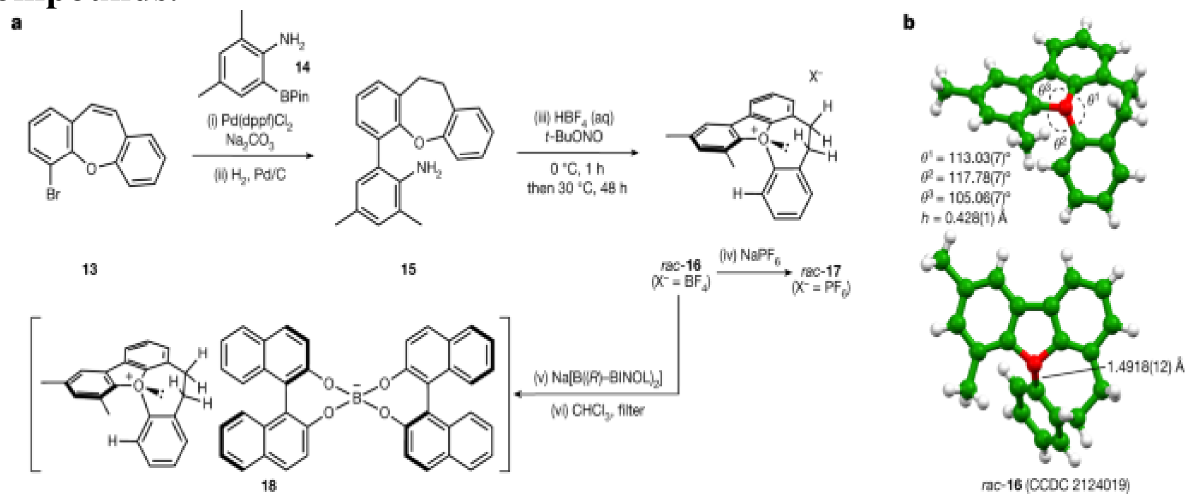
$sp^3$ carbon would actively lead to TS destabilization and a higher barrier to inversion.

We computed GS and TS structures for oxepin-containing oxonium ions (Fig. 2c(ii)). There is a substantial predicted increase in the barrier to inversion upon changing from the six to the seven-membered ring oxonium, from 59.0 to 117.0 kJ $mol^{-1}$ (for $R^2$ = Me, $R^3$ = H) and this is augmented to 127.2 kJ $mol^{-1}$ and 154.4 kJ $mol^{-1}$ upon introduction of larger *iso*-propyl and *tert*-butyl groups respectively. In the calculated dihydrodibenzooxepine GS structures, the sum of angles about oxygen is generally lower than in the xanthene series and decreases as the size of the substituent on the benzofuran ring increases; this is consistent with a greater apex height in the seven-ring series and greater pyramidalization about oxygen.

With this information in hand, we focussed on the synthesis of seven-ring-containing compounds, by using a strategically similar approach to that adopted earlier: dibenzooxepine **13** (made in four steps) was cross-coupled with aryl boronic ester **14** to afford a biaryl-containing intermediate in 67% yield. Hydrogenation gave **15**, which was diazotised under standard conditions and gently warmed to 30 °C to enable O-arylation and to afford the oxonium tetrafluoroborate *rac*-**16** in 75% yield (Fig. 3a). Counterion metathesis with sodium hexafluorophosphate enabled the synthesis of oxonium ion *rac*-**17** in 56% yield. Oxonium tetrafluoroborate *rac*-**16** and hexafluorophosphate *rac*-**17** are room-temperature stable solids that can be stored indefinitely without special precautions. The structure of *rac*-**16** was confirmed by single-crystal X-ray diffraction studies (space group $P-1$), which illustrate the expected puckered conformation of the oxepin ring and a pyramidal arrangement of atoms around the oxygen (the sum of the angles around oxygen is 335.9°, Fig. 3b). The apex height $h$ = 0.428(1) Å, reflects the increased pyramidalization about oxygen versus **10**. The external C–O–C angle is 117.78(7)° and the internal oxepin C–O–C angle is 113.03(7)°; these constitute relatively minor angular deviations from those observed in the parent heterocycles and are consistent with a ring system that does not experience substantial angle strain. Computed hyperhomodesmotic reaction enthalpies suggest that oxonium ion **16** is less strained than **10** by 13.4 kJ $mol^{-1}$. Examination of the unit cell does not show any short C–H or C–C distances less than normal van der Waals contact distances, which is

indicative that the conformation of the oxonium cation is not influenced by crystal packing. The longest C–O bond length is 1.4918(12) Å; this is longer than the corresponding measurement in benzofuran-derived triaryloxonium ions (1.459(6) Å) and the average C–O bond length in triphenyloxonium tetraphenylborate (at 1.47 Å).

**Fig. 3: Synthesis and structural study of stereogenic-at-oxygen compounds.**




**a**, Reagents and conditions: (i) Pd(dppf)$Cl_2$·$CH_2Cl_2$ (5 mol%), $Na_2CO_3$ (4 equivalents), PhMe/$H_2O$ (1:1, v/v), 95 °C, 24 h (67% yield); (ii) $H_2$ (g, 1 atmosphere), Pd/C (2.5 mol%), EtOAc, RT, 16 h (quantitative); (iii) $HBF_4$ (48% aq, 5 equvialents), *t*-BuONO (5 equivalents), $CH_2Cl_2$/IPA (1:1), 0 °C, 1 h then 30 °C, 48 h (75% yield); (iv) $NaPF_6$ (aq), $CH_2Cl_2$ (56% yield); (v) Na[B((*R*)−BINOL)$_2$] (1.0 equivalent), MeCN, RT then $CH_2Cl_2$, wash with water; (vi) dissolve in $CHCl_3$ and modulate concentration. **b**. Crystal and molecular structure of fused pentacyclic oxonium salt *rac*-**16** from single-crystal X-ray diffraction studies. The tetrafluoroborate counterion is omitted for clarity. Numbers in parentheses indicate standard uncertainties. *h,* height of conceptualized triangular oblique pyramid with oxygen at its apex and the adjacent carbons of the attached O-aryl rings forming the base.

Treatment of *rac*-**16** with a stoichiometric amount of sodium bis-(*R*)-BINOL borate led to complete conversion to a 1:1 mixture of diastereoisomeric salts (**18**, Fig. 3 and **19** (not shown)). Exploitation of the differential solubility of these two salts in chloroform enabled precipitation of one to yield the

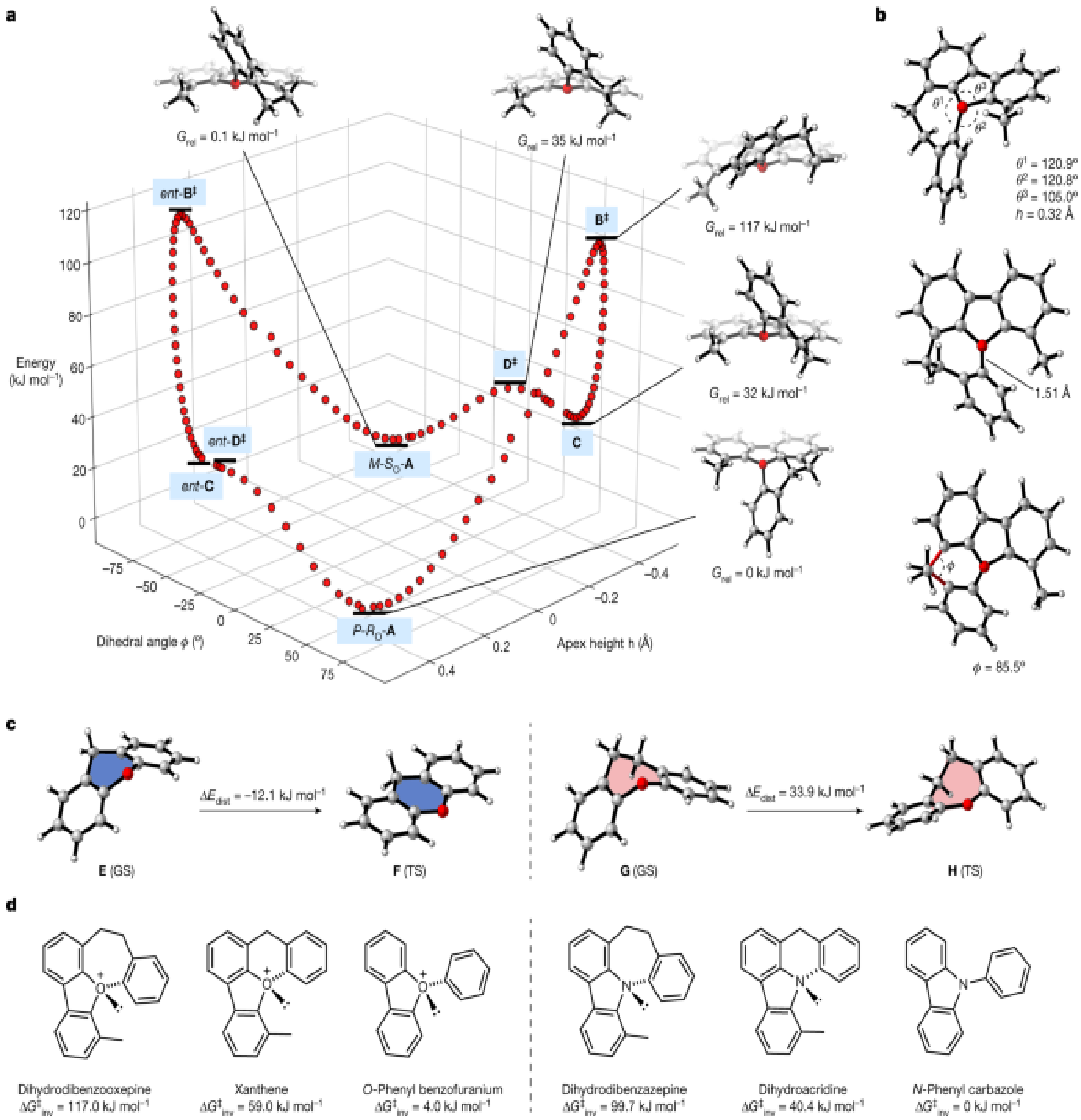
enriched diastereoisomeric pair **18** (ratio 94:6–99:1). We were unable to grow crystals suitable for X-ray analysis of diastereoisomerically enriched salt **18**, and hence we attempted to exchange the counterion in a highly diastereoisomerically enriched sample by stirring with an excess of tetrafluoroboric acid. From this mixture we were able to isolate crystals of the desired tetrafluoroborate salt **16**; this crystallized in $P2_12_12_1$, with an absolute configuration of $P$-$(R)_O$ (determined by refinement of the Flack parameter $x$ against X-ray diffraction datasets collected from three separate crystals; CCDC 21911038-2191040)[45], which thus also indicates the absolute configuration of the oxonium ion in **18** (see page 19 of the Supplementary Information). We probed the configurational stability of dichloromethane solutions of **18**, and observed a small change in the diastereoisomeric ratio over 7 days at 25 °C. With this data and the equilibrium ratio of 50.5:49.5 (**18**, $P$-$(R)_O$, ($R$); **19**, $M$-$(S)_O$,($R$)), we can calculate the barriers to helical inversion to be 111.8 ± 0.1 kJ mol$^{-1}$ (for the $P$-$(R)_O$, ($R$) **18** to $M$-$(S)_O$, ($R$) **19** conversion) and 111.7 ± 0.1 kJ mol$^{-1}$ (for the $M$-$(S)_O$, ($R$) **19** to $P$-$(R)_O$, ($R$) **18** conversion); attempts to examine equilibration at higher temperatures (more than 50 °C) led to decomposition at a rate faster than equilibration. We were also able to resolve oxonium hexafluorophosphate *rac*-**17** by chiral stationary phase high-performance liquid chromatography, and use this method to determine its enantiomerization barrier as 111.0 ± 0.1 kJ mol$^{-1}$. This demonstrates that the influence of the counterion on the magnitude of the inversion barrier is probably negligible, consistent with the observed values for diastereoisomeric salts **18** and **19** (which are almost energetically degenerate).

The experimentally determined barriers to inversion in **17** and **18** (111.0–111.8 kJ mol$^{-1}$) are close to that predicted by calculation (117.0 kJ mol$^{-1}$). A calculated reaction energy profile for the reversible enantiomerization of this oxonium ion is shown in Fig. 4a (ref. [46]); this indicates that the enantiomerization process from $P$-$(R)_O$ **A** consists of a high barrier helical inversion and a much lower barrier conformational relaxation associated with the seven-membered ring. The transition structure (TS) **B** for the helical inversion ($G_{rel}$ = 117 kJ mol$^{-1}$) possesses inverted pyramidal geometry at the oxygen, albeit reduced from the GS **A** ($h$ = 0.32 Å, sum of angles around

oxygen 346.7° for **B** versus 335.9° for **A**). A structure in which the oxygen is trigonal planar lies earlier on the reaction coordinate from **A** to **B**; this also demonstrates that O-inversion is directly coupled with overall inversion of the helical chirality of the molecule. In TS **B**, both the internal oxepin and external C–O–C angles are close to 121° (Fig. 4b), which represents a notable change in structure around the oxygen atom versus the GS. Lengthening of the C–O bond in the oxepin ring versus the GS is also observed (1.51 versus 1.49 Å). TS **B** leads directly to intermediate **C** ($G_{rel}$ = 32 kJ mol$^{-1}$) in which the oxepin ring populates a boat-like conformation; this undergoes a facile ring flip via TS **D** ($G_{rel}$ = 35 kJ mol$^{-1}$) to the more stable boat-like conformation in GS structure $M$-$(S)_O$ **A**. To rationalize why the inversion barrier of the oxepin-containing oxonium ion is greater than 40 kJ mol$^{-1}$ higher than its xanthene-containing analogue, we calculated how ring strain in the six- and seven-membered rings in oxonium ions changes on going from the GS to the TS (Fig. 4c). In the xanthene-containing system, the boat-like GS **E** changes to an almost perfectly planar TS **F** through movement of the two aromatic rings into the same plane. This leads to a decrease in strain in the six-membered heterocycle of 12.1 kJ mol$^{-1}$. This is owing to the relief of angle strain in the xanthene ring, which can be seen in the increase in internal C–O–C angle from 111.0° to 116.6° (GS to TS). In this system, the TS resembles the GS conformation of xanthene itself. In contrast, there is an increase in strain of 33.9 kJ mol$^{-1}$ in the dihydrodibenzooxepine ring going from GS **G** to TS **H**. To attain the TS geometry demonstrated in **H**, the two aromatic rings twist to enable the inversion of the seven-membered ring. This raises the energy through a combination of torsional strain (between the two sp$^3$-carbon atoms in the backbone (H–C–C–H torsion = 38.1°)), an increase in allylic strain between the backbone sp$^3$ C–H bonds and the aromatic rings, internal angle strain and a minor deviation from planarity in the two aromatic rings. The seven-membered ring conformation in oxonium ion **G** is boat-like and resembles dihydrodibenzooxepine itself: deviation from this geometry in TS **H** raises the energy substantially. Delocalization of the oxygen lone pair changes along the reaction coordinate for inversion and this is also a prominent source of differentiation between dihydrodibenzooxepine and xanthene-derived oxonium ions. Natural bond orbital[47] calculations highlight how oxygen lone-pair delocalization into the adjacent π systems stabilizes the

inversion TS. For the xanthene oxonium ion, the increasing planarity about oxygen stabilizes the TS by 54.1 kJ mol$^{-1}$ relative to the GS, whereas for the dihydrodibenzooxepine oxonium ion this is lower at 12.0 kJ mol$^{-1}$, a difference of 33.1 kJ mol$^{-1}$. Increased strain in the seven-membered ring oxonium backbone resists planarization in the inversion TS, reducing the extent of lone-pair delocalization, which consequently increases the activation energy barrier.

**Fig. 4: Mechanistic overview of enantiomerization of oxepin-containing oxonium ion.**



**a**, B3LYP-D3(BJ)/6-31+G(d,p) computed Gibbs energy profile for enantiomerization demonstrates that the process consists of a high barrier oxygen inversion and a low barrier conformational relaxation of the seven-membered ring. **b**, Geometrical measurements of TS **B**. $\phi$ = (C6–C7–C8–C9) dihedral angle. **c**, Computed change in ring strain ($\Delta E_{dist}$) in highlighted rings in xanthene and dihydrodibenzooxepine oxonium ions on going from GS to TS. The relative energies of truncated xanthene and dihydrodibenzooxepine substructures are computed at geometries taken from the parent oxonium ion. **d**, Comparison of computed barriers to inversion show that triaryloxonium ions have higher barriers to inversion than corresponding triaryl amines (B3LYP-D3(BJ)/6-31+G(d,p)). $G_{rel}$ = relative Gibbs free energy. $^{‡}$ = transition state.

To examine how the identity of the heteroatom influences the inversion barrier, we studied this class of oxonium ions and their nitrogen analogues computationally (Fig. 4d). The calculated barriers to inversion of triaryloxonium ions are systematically higher than their triarylamine counterparts. This can be rationalized based on the resistance to rehybridization of the heteroatom, which is greater in the oxonium ion (see page 38 of the Supplementary Information). For the nitrogen analogues there is greater lone-pair delocalization in the inversion TS structures. This is consistent with calculations on the dihydrodibenzooxepine oxonium ($\Delta G^{‡}_{inv}$ = 117.0 kJ mol$^{-1}$) and the dihydrodibenzazepine ($\Delta G^{‡}_{inv}$ = 99.7 kJ mol$^{-1}$) systems, indicating that barrier height depends on both the geometric restraint provided by the ring structure and the identity of the heteroatom.

# Data availability

All data (experimental procedures, characterization data and cartesian coordinates for all DFT calculations) supporting the findings of this study are available within the article and its supplementary information. Crystallographic data for compounds **10**, *rac*-**16** and *P*-$(R)_O$-**16** (three datasets) have been deposited with the Cambridge Crystallographic data Centre under deposition numbers CCDC 2124020, CCDC 2124019 and CCDC 21911038–2191040. These data can be obtained free of charge from www.ccdc.cam.ac.uk/data_request/cif.

# References


1. Lehn, J. M. Nitrogen inversion: experiment and theory. *Fortschr. Chem. Forsch.* **15**, 311–377 (1970).

2. Baechler, R. D. & Mislow, K. Effect of structure on the rate of pyramidal inversion of acyclic phosphines. *J. Am. Chem. Soc.* **92**, 3090–3093 (1970).

3. Andersen, K. K., Cinquini, M. & Papanikolaou, N. E. Synthesis and stereochemistry of triarylsulfonium salts. *J. Org. Chem.* **35**, 706–710 (1970).

4. Rauk, A., Allen, L. C. & Mislow, K. Pyramidal inversion. *Angew. Chem. Int. Ed. Engl.* **9**, 400–414 (1970).

5. Chan, H. S., Nguyen, Q. N. N., Paton, R. S. & Burton, J. W. Synthesis, characterization and reactivity of complex oxonium ions, proposed intermediates in natural product biosynthesis. *J. Am. Chem. Soc.* **141**, 15951–15962 (2019).

6. Meerwein, H., Hinz, G., Hofmann, P., Kroning, E. & Pfeil, E. Über tertiäre oxoniumsalze I. *J. Prakt. Chem.* **147**, 257–285 (1937).

7. Lambert, J. B. & Johnson, D. H. The stereochemistry and inversion of trivalent oxygen. *J. Am. Chem. Soc.* **90**, 1349–1350 (1968).

8. Tröger, J. Ueber einige mittelst nascirenden formaldehydes entstehende basen. *J. Prakt. Chem.* **36**, 225–245 (1887).

9. Prelog, V. & Wieland, P. Über die spaltung der Tröger'schen base in optische antipoden, ein beitrag zur stereochemie des dreiwertigen stickstoffs. *Helv. Chim. Acta* **27**, 1127–1134 (1944).

10. Rigol, S., Beyer, L., Hennig, L., Sieler, J. & Giannis, A. Hunlich base: (re)discovery, synthesis, and structure elucidation after a century. *Org. Lett.* **15**, 1418–1420 (2013).

11. Boratyński P.J., Zielińska-Błajet M. & Skarżewski J. Cinchona alkaloids-derivatives and Applications. *Alkaloids Chem. Biol.* **82**, 29−145 (2019).

12. Huang, S. et al. Organocatalytic enantioselective construction of chiral azepine skeleton bearing multiple-stereogenic elements. *Angew. Chem. Int. Ed. Engl.* **60**, 21486–21493 (2021).

13. Rowley, J. H., Yau, S. C., Kariuki, B. M., Kennedy, A. R. & Tomkinson, N. C. O. Readily accessible chiral at nitrogen cage structures. *Org. Biomol. Chem.* **11**, 2198–2205 (2013).

14. Walsh, M. P., Phelps, J. M., Lennon, M. E., Yufit, D. S. & Kitching, M. O. Enantioselective synthesis of ammonium cations. *Nature* **597**, 70–76 (2021).

15. Koeppl, G. W., Sagatys, D. S., Krishnamurthy, G. S. & Miller, S. I. Inversion barriers of pyramidal ($XY_3$) and related planar (=XY) species. *J. Am. Chem. Soc.* **89**, 3396–34051 (1967).

16. Penner, G. H., Polson, K. M., Daleman, S. & Reid, K. Deuterium NMR study of molecular dynamics and geometry in two classes of onium salts: $(CH_3)E^+X^-$ and $C_6H_5M(CH_3)_3{}^+I^-$. *Can. J. Chem.* **71**, 417–426 (1993).

17. Mikata, Y., Fujimoto, T., Noko, I. & Kondo, S. Differentiation of oxygen atom chirality in copper(II) complexes with dipicolylamine-derived ligands.*Eur. J. Inorg. Chem.* **2012**, 4310–4317 (2012).

18. Mikata, Y., Sugai, Y., Obata, M., Harada, M. & Yano, S. Novel oxygen chirality induced by asymmetric coordination of an ether oxygen atom to a metal center in a series of sugar-pendant dipicolylamine copper(II) complexes. *Inorg. Chem.* **45**, 1543–1551 (2006).

19. Braddock, D. C. A hypothesis concerning the biosynthesis of the obtusallene family of marine natural products via electrophilic bromination. *Org. Lett.* **8**, 6055–6058 (2006).

20. Murai, A. in *Comprehensive Natural Products Chemistry* 1st edn (eds Barton, D. H. R., Meth-Cohn & O. Nakanishi, K.) 303−324 (Elsevier, 1999).

21. Kikuchi, H., Suzuki, T., Kurosawa, E. & Suzuki, M. The structure of notoryne, a halogenated $C_{15}$ nonterpenoid with a novel carbon skeleton from the red alga *Laurenda nipponica* yamada. *Bull. Chem. Soc. Jpn.* **64**, 1763–1775 (1991).

22. Bonney, K. J. & Braddock, D. C. A unifying stereochemical analysis for the formation of halogenated $C_{15}$-acetogenin medium-ring ethers from *laurencia* species via intramolecular bromonium ion assisted epoxide ring-opening and experimental corroboration with a model epoxide. *J. Org. Chem.* **77**, 9574–9584 (2012).

23. Kim, B. et al. Biomimetic asymmetric total synthesis of (−)-laurefucin via an organoselenium-mediated intramolecular hydroxyetherification. *J. Am. Chem. Soc.* **130**, 16807–16811 (2008).

24. Zhang, Y. A., Yaw, N. & Snyder, S. A. General synthetic approach for the *laurencia* family of natural products empowered by a potentially biomimetic ring expansion. *J. Am. Chem. Soc.* **141**, 7776–7788 (2019).

25. Stoyanov, E. S. et al. The $R_3O^+\cdots H^+$ hydrogen bond: toward a tetracoordinate oxadionium(2+) ion. *J. Am. Chem. Soc.* **134**, 707–714 (2012).

26. Meerwein, H., Bettenberg, E., Gold, H., Pfeil, E. & Willfang, G. Über tertiäre oxoniumsalze II. *J. Prakt. Chem.* **154**, 83–156 (1940).

27. Klages, F. & Jung, H. A. Synthese und eigenschaften eines bicyclischen oxoniumsalzes. *Chem. Ber.* **98**, 3757–3762 (1965).

28. Gargulio, R. J. & Tarbell, S. Cyclic oxonium ions and related structures. *Proc. Natl Acad. Sci. USA.* **62**, 52–55 (1969).

29. Olah, G. A., Doggweiler, H., Felberg, J. D. & Frohlich, S. Onium ions. 33. (Trimethylsilyl)- and [(trimethylsilyl)methyl]oxonium and -

halonium ions. *J. Org. Chem.* **50**, 4847–4851 (1985).

30. Etzhorn, M.et al. 1-Oxoniaadamantane. *Eur. J. Org. Chem.* **2008**, 4555–4558 (2008).

31. Mascal, M., Hafezi, N., Meher, N. K. & Fettinger, J. C. Oxatriquinane and oxatriquinacene: extraordinary oxonium ions. *J. Am. Chem. Soc.* **130**, 13532–13533 (2008).

32. Mascal, M., Hafezi, N. & Toney, M. D. 1,4,7-Trimethyloxatriquinane: $S_N2$ reaction at tertiary carbon. *J. Am. Chem. Soc.* **132**, 10662–10664 (2010).

33. Gunbas, G. et al. Extreme oxatriquinanes and a record C-O bond length. *Nat. Chem.* **4**, 1018–1023 (2012).

34. Suzuki, H. & Muratake, H. Functionalized oxatriquinanes and their structural equilibrium in protic solvent. *Chem. Pharm. Bull.* **62**, 921–926 (2014).

35. Nesmeyanov, A. N., Makarova, L. G. & Tolstaya, T. P. Heterolytic decomposition of onium compounds (diphenyl halogenonium and triphenyloxonium salts). *Tetrahedron* **1**, 145–157 (1957).

36. Watkins, M. I., Ip, W. M., Olah, G. A. & Bau, R. Onium ions. 23. Structure of oxonium ions: an X-ray crystallographic study of triethyloxonium hexafluorophosphate and triphenyloxonium tetraphenylborate. *J. Am. Chem. Soc.* **104**, 2365–2372 (1982).

37. Tolstoya, T. P. et al. Nucleophilic substitution in *O*-phenyldibenzofuranium and 10-phenylxanthonium cations. *Russ. Chem. Bull.* **46**, 789–797 (1997).

38. Hellwinkel, D. & Seifert, H. Ringschlußreaktionen von 2′-heterosubstituierten biphenyl-2-diazonium-salzen zu (spiro)cyclischen tetraarylammonium-salzen und tribenz[b.d.f]azepinen. *Chem. Ber.* **105**, 880–906 (1972).

39. Oki, M. in *Topics in Stereochemistry* Vol. 14 (eds Allinger, N. L., Eliel, E. L. & Wilen, S. H.) 1–81 (Wiley, 1983).

40. Lu, M. et al. Peraryl-x-onium ions of nitrogen and oxygen. *Org. Chem. Front.* **6**, 2640–2646 (2019).

41. Bosson, J., Gouin, J. & Lacour, J. Cationic triangulenes and helicenes: synthesis, chemical stability, optical properties and extended applications of these unusual dyes. *Chem. Soc. Rev.* **43**, 2824–2840 (2014).

42. Wu, Y.-T., Hayama, T., Baldridge, K. K., Linden, A. & Siegel, J. S. Synthesis of fluoranthenes and indenocorannulenes: elucidation of chiral stereoisomers on the basis of static molecular bowls. *J. Am. Chem. Soc.* **128**, 6870–6884 (2006).

43. Zhukhlistova, N. E., Tishchenko, G. N., Tsarev, D. A. & Tolstaya, T. P. Crystal and molecular structure of *O*-phenyldibenzofuranium tetrafluoroborate. *Kristallografiya* **42**, 863–869 (1997).

44. Wheeler, S. E., Houk, K. N., Schleyer, P. V. R. & Allan, W. D. A hierarchy of homodesmotic reactions for thermochemistry. *J. Am. Chem. Soc.* **131**, 2547–2560 (2009).

45. Flack, H. D. On enantiomorph-polarity estimation. *Acta Crystallogr.* **A39**, 876–881 (1983).

46. Reist, M., Testa, B., Carrupt, P. A., Jung, M. & Schurig, V. Racemization, enantiomerization, diastereomerization and epimerization: their meaning and pharmacological significance. *Chirality* **7**, 396–400 (1995).

47. Glendening, E. D. et al. NBO v. 7.0. Theoretical Chemistry Institute, University of Wisconsin, Madison https://nbo6.chem.wisc.edu/biblio_css.htm (2018).

# Acknowledgements

The EPSRC has provided financial support for studentships (to O.S. and M.V.P.) via the Centre for Doctoral Training in Synthesis for Biology and Medicine (EP/ L015838/1) and to M.J.H. via a DTP studentship (EP/T517811/1; project no. 2446204). R.P. acknowledges support from the NSF (CHE-1955876) and computational resources from the RMACC Summit supercomputer supported by the National Science Foundation (ACI-1532235 and ACI-1532236), the University of Colorado Boulder and Colorado State University and the Extreme Science and Engineering Discovery Environment (XSEDE) through allocation TG-CHE180056. We are grateful to A. Thompson (University of Oxford) for assistance with X-ray crystallography and P. Solon for assistance with the high-performance liquid chromatography measurements. A CC-BY licence is applied to the author-accepted manuscript arising from this submission, in accordance with open-access conditions of the EPSRC.

# Author information

## Authors and Affiliations

1. Chemistry Research Laboratory, University of Oxford, Oxford, UK

   Owen Smith, Mihai V. Popescu, Madeleine J. Hindson, Jonathan W. Burton & Martin D. Smith

2. Department of Chemistry, Colorado State University, Ft. Collins, CO, USA

   Mihai V. Popescu & Robert S. Paton

## Contributions

O.S., M.V.P., M.J.H., R.P., J.W.B. and M.D.S. conceived and designed the study. O.S. and M.J.H. performed the synthetic experiments and analysed data for all compounds. O.S. performed the X-ray crystallography. M.V.P. and R.P. performed the computational study. O.S., M.V.P., M.J.H., R.P., J.W.B. and M.D.S. cowrote the paper. Correspondence regarding

computation should be addressed to R.P.; all other correspondence should be addressed to M.D.S. and J.W.B.


### Corresponding authors

Correspondence to Robert S. Paton, Jonathan W. Burton or Martin D. Smith.


## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Peer review

### Peer review information


*Nature* thanks Jonathan Clayden, Ganna Gryn'ova and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.


## Additional information


**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.


## Supplementary information

### Supplementary Information

### Crystallographic data.

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

# About this article

## Cite this article

Smith, O., Popescu, M.V., Hindson, M.J. *et al.* Control of stereogenic oxygen in a helically chiral oxonium ion. *Nature* **615**, 430–435 (2023). https://doi.org/10.1038/s41586-023-05719-z

- Received: 26 July 2022
- Accepted: 09 January 2023
- Published: 15 March 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05719-z

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

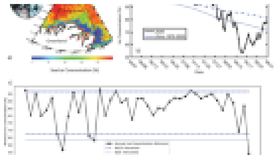
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

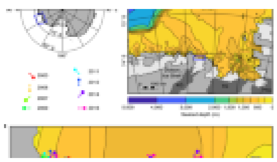
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05719-z



## Accelerated sea ice loss in the Wandel Sea points to a change in the Arctic’s Last Ice Area

01 July 2021

Axel J. Schweiger, Michael Steele, … Kristin L. Laidre



## West Antarctic Ice Sheet retreat in the Amundsen Sea driven by decadal oceanic variability

13 August 2018

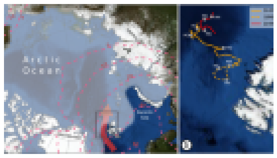
Adrian Jenkins, Deb Shoosmith, … Sharon Stammerjohn



## Natural drivers of multidecadal Arctic sea ice variability over the last millennium

20 January 2020

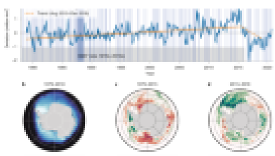
Paul R. Halloran, Ian R. Hall, … Freya Garry



## [Winter storms accelerate the demise of sea ice in the Atlantic sector of the Arctic Ocean]

25 June 2019

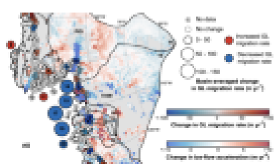
Robert M. Graham, Polona Itkin, … Mats A. Granskog



## [Rapid decline in Antarctic sea ice in recent years hints at future change]

28 June 2021

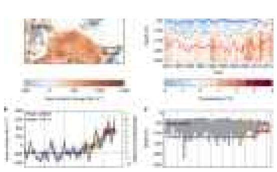
Clare Eayrs, Xichen Li, … David M. Holland



## [Inter-decadal climate variability induces differential ice response along Pacific-facing West Antarctica]

16 January 2023

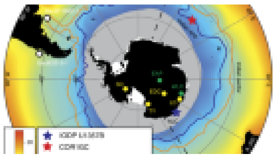
Frazer D. W. Christie, Eric J. Steig, … Robert G. Bingham



## [Arctic warming hotspot in the northern Barents Sea linked to declining sea-ice import]

25 June 2018

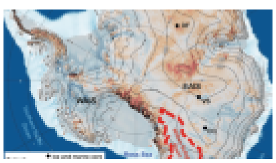
Sigrid Lind, Randi B. Ingvaldsen & Tore Furevik



## [Multi-decadal trends in Antarctic sea-ice extent driven by ENSO–SAM over the last 2,000 years]

22 February 2021

Xavier Crosta, Johan Etourneau, … Minoru Ikehara



## [Wilkes subglacial basin ice sheet response to Southern Ocean warming during late Pleistocene interglacials]

10 September 2022

Ilaria Crotti, Aurélien Quiquet, … Massimo Frezzotti

- Article
- Open Access
- Published: 15 March 2023

# Regime shift in Arctic Ocean sea ice thickness


- Hiroshi Sumata ,
- Laura de Steur,
- Dmitry V. Divine,
- Mats A. Granskog &
- …
- Sebastian Gerland

Show authors



*Nature* **volume 615**, pages 443–449 (2023)



- 4277 Accesses
- 935 Altmetric
- Metrics details


## Abstract


Manifestations of climate change are often shown as gradual changes in physical or biogeochemical properties[1]. Components of the climate system, however, can show stepwise shifts from one regime to another, as a nonlinear response of the system to a changing forcing[2]. Here we show that the Arctic sea ice regime shifted in 2007 from thicker and deformed to thinner and more uniform ice cover. Continuous sea ice monitoring in the Fram Strait over the last three decades revealed the shift. After the shift, the fraction of thick and deformed ice dropped by half and has not recovered to date. The timing of the shift was preceded by a two-step reduction in residence time of sea ice in the Arctic Basin, initiated first in 2005 and followed by 2007. We demonstrate that a simple model describing the stochastic process of dynamic sea ice thickening explains the observed ice thickness changes as a result of the reduced residence time. Our study

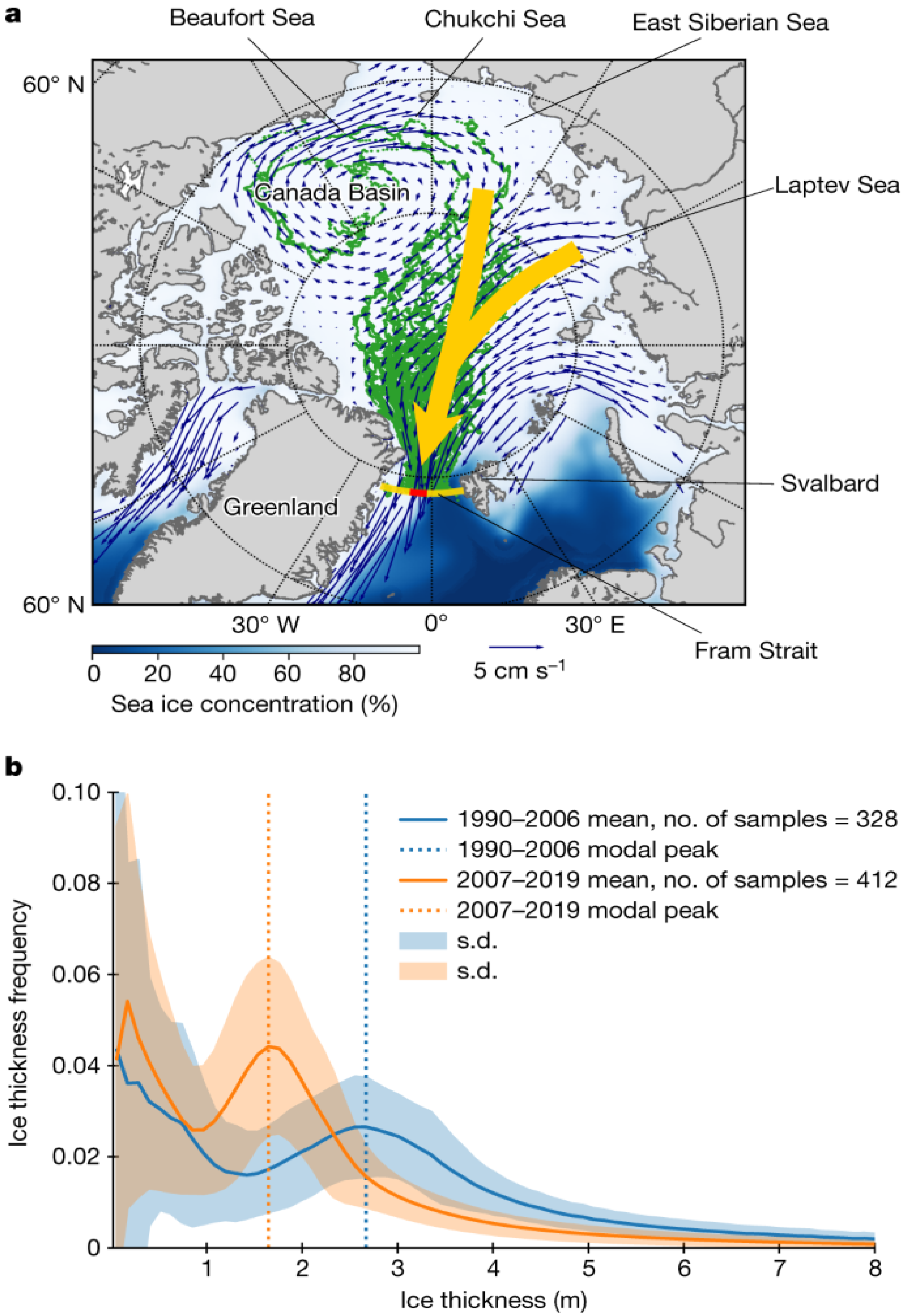
highlights the long-lasting impact of climate change on the Arctic sea ice through reduced residence time and its connection to the coupled ocean–sea ice processes in the adjacent marginal seas and shelves of the Arctic Ocean.

# Main

The extent of Arctic sea ice exhibited negative trends both in summer and winter over the last three decades[1,3]. Such a retreat in summer sea ice results in an increase of open water areas and prolongation of ice-free conditions in the adjacent marginal seas and shelves of the Arctic Ocean[4,5]. The adjacent, ice-free seas absorb more solar energy during summer and store heat in the upper ocean[6]. The heat enhances ice melt in the marginal ice zone in summer and delays the onset of new ice formation in the autumn[7]. These processes influence Arctic-wide sea ice properties through large-scale sea ice motion carrying the ice from the marginal seas to the central Arctic, a process known as Transpolar Drift (TPD) Stream[8].

A large fraction of the Arctic Ocean's sea ice is formed in the marginal seas of the Arctic Ocean, from the Alaskan to Siberian sectors of the Arctic[9], that is, the Beaufort, Chukchi, East Siberian and Laptev seas. It is then transported across the Arctic to sustain the perennial ice pack (Fig. 1a). A main part of sea ice formed in the Alaskan sector circulates in the Canada Basin[10], while a part of the ice is pushed towards the Siberian sector via the Chukchi Sea[11]. Sea ice formed in the Siberian sector, together with the ice from the Alaskan sector, joins the TPD emanating from offshore of the Siberian shelves. The TPD transports the ice to the central Arctic and further towards the Atlantic sector of the Arctic[8,12]. Finally, it exits the Arctic Ocean through the Fram Strait, located between north-eastern Greenland and Svalbard. Sea ice motion and ice age estimates from observations indicate an acceleration of ice motion and hence decrease of residence time of sea ice in the Arctic Basin in recent decades[13,14].

**Fig. 1: Map of Arctic Ocean and sea ice thickness distribution in the Fram Strait.**


a
Beaufort Sea
Chukchi Sea
East Siberian Sea
Laptev Sea
Canada Basin
Svalbard
Greenland
Fram Strait
60° N
60° N
30° W
0°
30° E
0
20
40
60
80
Sea ice concentration (%)
5 cm s−1
b
1990–2006 mean, no. of samples = 328
1990–2006 modal peak
2007–2019 mean, no. of samples = 412
2007–2019 modal peak
s.d.
s.d.
Ice thickness frequency
0
0.02
0.04
0.06
0.08
0.10
1
2
3
4
5
6
7
8
Ice thickness (m)

**a**, Arctic Ocean and its marginal seas, with winter sea ice concentration (1980–2018 mean, white-blue shading, calculated from OSI SAF[51]), ice drift field (blue arrows, Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors v.4.1)[52], 83 ice-tethered buoy tracks that arrived in the Fram Strait (green lines) and TPD Stream (yellow shade). The buoy tracks were obtained from the International Arctic Buoy Programme[53]. The Fram Strait Arctic Outflow Observatory is shown by the red bar. **b**, Mean sea ice thickness distribution in the Fram Strait before and after 2007. The distributions were derived on a monthly basis by all available ULS data from 1990 to 2019 (described in the Methods) and averaged across two periods: 1990–2006 and 2007–2019. The Matplotlib basemap toolkit was used to plot the map.

Because up to approximately 90% of sea ice outflow from the Arctic Ocean to the Subarctic North Atlantic occurs through the Fram Strait[15,16], sea ice properties observed in the Fram Strait represent basin-wide characteristics of Arctic sea ice[14,17]. The Fram Strait Arctic Outflow Observatory has been monitoring sea ice and ocean properties in the core of the outflow at a latitude of approximately 79° N[18] (marked red in Fig. 1a) since 1990. The observatory has provided a unique, near-continuous time series of sea ice thickness for the last three decades[19] (Methods).

In this study, we show that in 2007, a regime shift of sea ice thickness occurred in the Fram Strait. The shift was characterized by an abrupt reduction of deformed thick ice (a 52% reduction of sea ice thicker than 4 m) and an increased uniformity of ice thickness distribution (height of the modal peak increased by 67%). The timing of the shift follows a drop in Arctic-wide sea ice residence time in 2005 and 2007. We introduce a simple model describing the stochastic process of dynamic ice thickening and explain the relationship between the changes in Arctic sea ice residence time and thickness distribution.

## Changes in ice thickness distribution

Sea ice observed in the Fram Strait consists of a variety of ice types and thicknesses, reflecting its thermal and dynamic history before exiting the

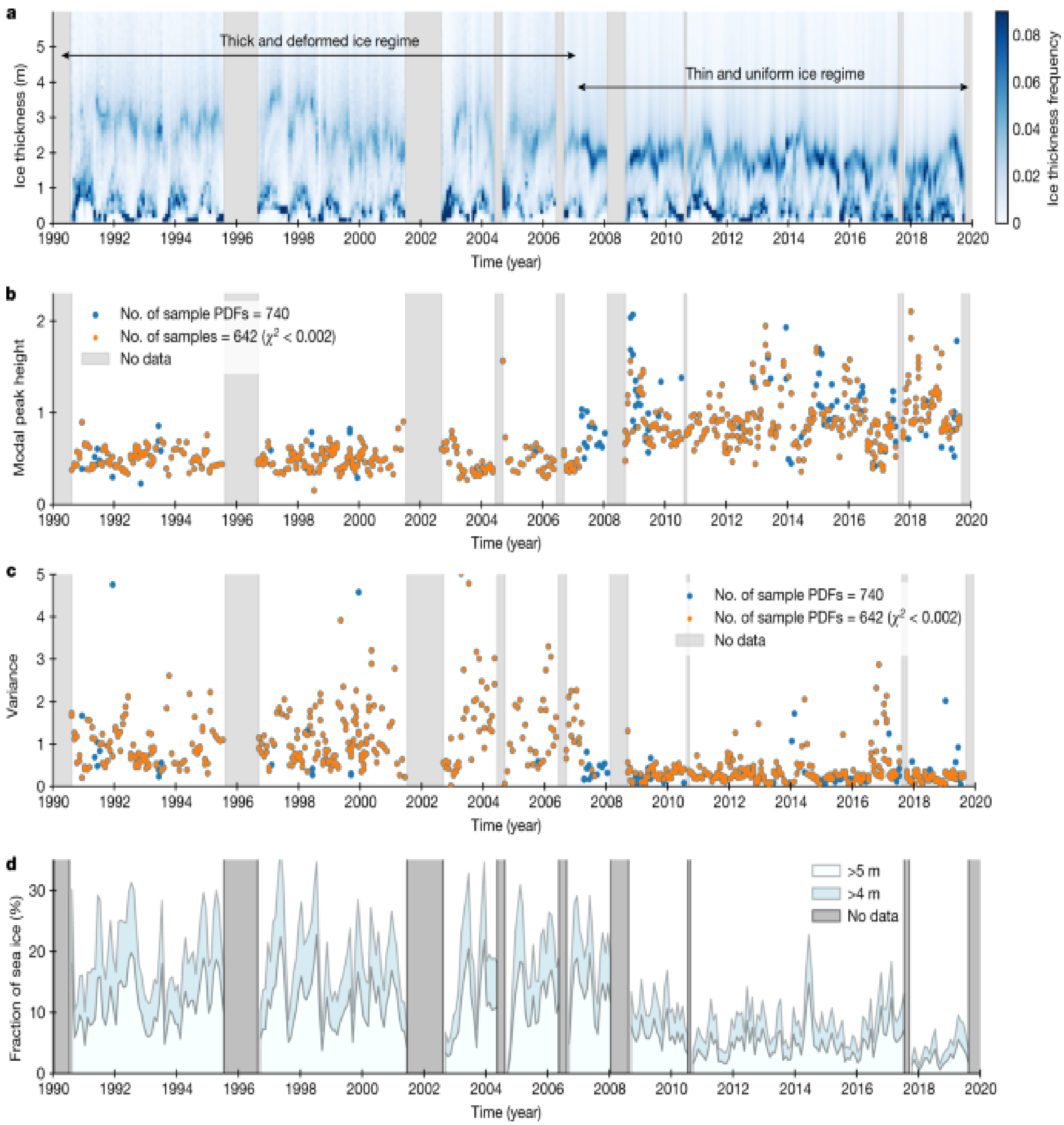
Arctic Basin[20,21]. Typically, ice thickness distribution in the Fram Strait is bimodal (Fig. 1b and Extended Data Fig. 5). The first peak represents thin sea ice (less than 0.5 m) formed in the vicinity of the Fram Strait, while the second peak (that is, the modal peak), which has a 1.5–3 m thickness, primarily represents thermodynamically grown sea ice that has been transported across the Arctic Basin. Finally, the distribution has a long tail towards thicker ice fractions, corresponding to dynamically thickened sea ice (from thinner thermodynamically grown sea ice) due to ridging and rafting. The second peak and tail can be reasonably well approximated using a log-normal function (Fig. 1b, Extended Data Fig. 5 and Methods). This part of the distribution represents sea ice properties in the central Arctic and gives information on combined thermal and dynamic forcing on sea ice from the time of its formation to its arrival in the Fram Strait.

The distribution of sea ice thickness has changed substantially in the last three decades, reflecting the environmental changes in the Arctic and consequent reduction of mean ice thickness in the Arctic Ocean[21]. In the Fram Strait, the thickness of the modal peak has been reduced by approximately 1 m (2.7 m to 1.7 m; Fig. 1b, dashed lines); a fraction of the ice in the mode (height of the peak) has increased by 67% from 1990–2006 to 2007–2019 (Fig. 1b, blue versus orange line). The tail of the distribution, corresponding to the deformed fraction of ice, has also substantially changed between the two periods: the thickness distribution for 1990–2006 (Fig. 1b, blue line) was characterized by thicker modal thickness with a smaller and broader modal peak and a larger fraction of deformed ice (thick and deformed ice regime), while the period after 2007 (Fig. 1b, orange line) is characterized by a thinner modal thickness with more compact distribution of ice thickness around the mode and a smaller fraction of deformed ice (thin and more uniform ice regime). These findings from the mooring observations are consistent with in situ observations[22].

Figure 2a shows the time series of ice thickness distribution in the Fram Strait for the last three decades. Darker shading indicates a larger fraction of sea ice at the corresponding thickness. A zonal band with a darker shade ranging approximately from 3.0 m to 1.5 m, depicts the modal thickness of multi-year sea ice. The temporal variation of this band describes the long-term changes of modal sea ice thickness in the central Arctic. The time

series clearly shows that the change of the thickness distribution (Fig. 1b) has not been a gradual process but that a distinct shift from thick and deformed ice regime to thinner and more uniform ice regime occurred around 2007 (visualized by the intensity of the shade at the modal peak in Fig. 2a). The shift is further evidenced by coincident changes in modal peak height (Fig. 2b) and the variance of ice thickness distribution (Fig. 2c). The modal peak height and variance were obtained from a fitted log-normal function to each distribution (Methods). The modal peak height gives a measure of compactness of the distribution around the mode, while the variance gives an indication of deformed fraction of sea ice relative to the modal peak. The sequential *t*-test analysis of regime shifts[23] we applied to these time series detected a shift in both peak height and variance in 2007 (Extended Data Fig. 6 and Extended Data Table 1). This indicates that until 2007, ice floes consisted of sea ice with a variety of thicknesses towards thicker ice, whereas from 2007, they have consisted of ice of more uniform thickness with a smaller deformed fraction. This is also clearly visible in shifts with a 1-year delay in the fraction of thick ice (that is, ice thickness exceeding 4 and 5 m) shown in Fig. 2d. This fraction corresponds to ice thicker than the maximum theoretical thickness of thermodynamically grown ice derived by different sea ice models[24,25], that is, giving a measure of the fraction of dynamically deformed ice. The shift of ice thickness distribution in 2007 and 2008 indicates that a sudden reduction of dynamic forcing on the ice occurred at that time.

**Fig. 2: Sea ice thickness properties observed in the Fram Strait in the last three decades.**

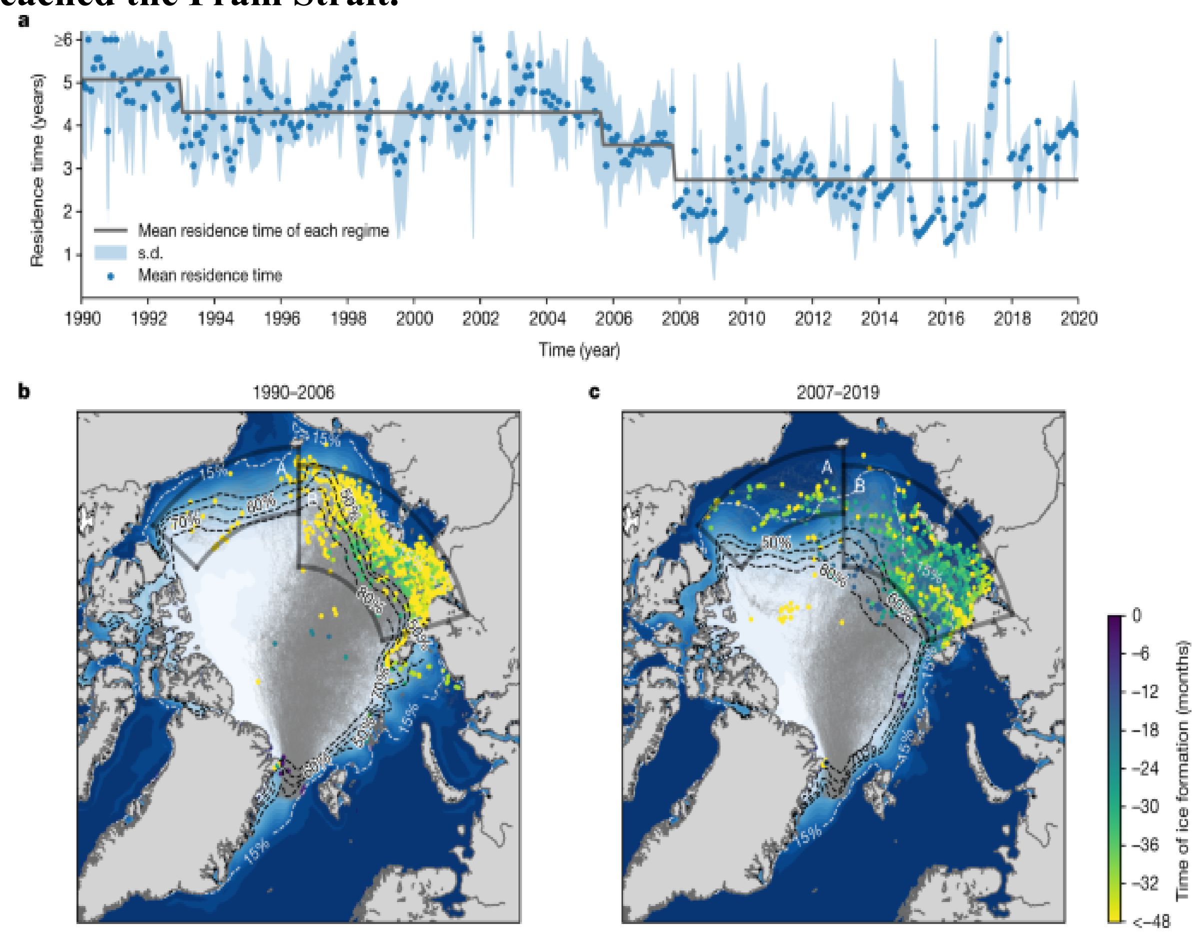



**a**–**d**, Time series of sea ice thickness distribution (**a**), modal peak height (**b**) and variance of ice thickness distributions (**c**), and fraction of sea ice thicker than two thresholds, that is, 4 m and 5 m, respectively (**d**). In **b**,**c**, $\chi^2$ is a sum of the squared residuals at each log-normal function fitting. Derivations of the ice thickness distribution, modal peak height and variance are described in the Methods.

# Reduced residence time of Arctic sea ice

Figure 3a shows a time series of mean residence time of sea ice floes in the Arctic Ocean that arrived in the Fram Strait (Methods). The timing of the observed shift in thickness distribution coincided with the timing of a reduction in residence time. The mean residence time showed a two-step shift from 4.3 years to 2.7 years in 2005 and 2007 (detected by sequential *t*-test analysis of regime shifts; Methods). A coincident exceptional long residence time (Fig. 3a) and large variance of thickness distribution (Fig. 2c) occurred in 2017, reinforcing the connection between residence time and thickness distribution. Although the mean residence time dropped in 2005 and 2007, the area of ice formation (coloured dots in Fig. 3b,c) and ice trajectories across the Arctic basins (grey cloud in Fig. 3b,c) have not changed notably. The slight offshore and westward shift of ice formation areas[26] is not sufficient to explain the reduction of residence time by approximately 1.6 years (Fig. 3a and colour of the dots in Fig. 3b,c).

**Fig. 3: Residence time and origins of sea ice in the Arctic Ocean that reached the Fram Strait.**

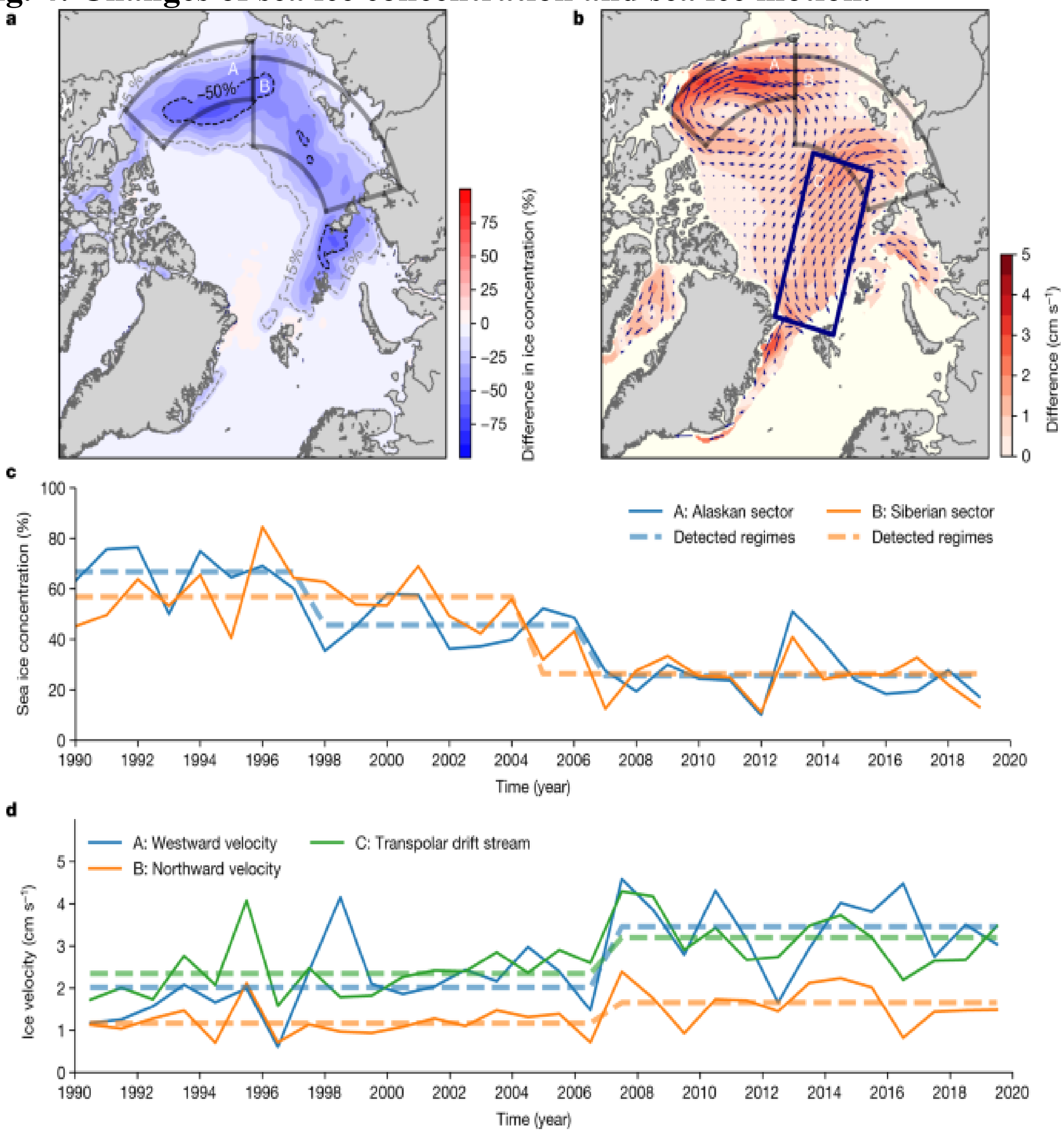


**a**–**c**, Time series of residence time of ice floes in the Arctic Basin (**a**) and origins and pathways of ice floes before (**b**) and after 2007 (**c**). **a**, The abscissa references the time of arrival in the Fram Strait. The grey line shows the mean residence time in each regime detected by sequential *t*-test analysis of regime shifts. **b**,**c**, The location of the dots depicts areas of sea ice formation, while the colour of the dots indicates the time of sea ice formation relative to their arrival in the Fram Strait. The grey clouds in **b** and **c** show the trajectories of ice floes from their origins to the Fram Strait. The background colour (navy–white shading) in **b** and **c** shows the mean sea ice concentration in September for the corresponding periods (OSI SAF[51]). The contours of the mean sea ice concentration are represented by the dashed lines (80%, 70%, 60% and 50% contours shown in black and 15% contour shown in white). Two polygons indicate the sea ice formation area in the Alaskan (A) and Siberian (B) sectors: the time series of sea ice concentration and ice drift speed in these areas are shown in Fig. 4. See the Methods for details of the backward trajectory calculation and residence time estimates. The Matplotlib basemap toolkit was used to plot the map.

The reduction of summer sea ice concentration in areas of sea ice formation, on the other hand, correlates with the reduction in residence time ($r = 0.65$ in the Alaskan sector, 0.73 in the Siberian sector, ice concentration leads 1 year). Figure 4a shows the difference in September sea ice concentration between two periods, 1990–2006 and 2007–2019. Areas that show the largest decrease coincide with areas of sea ice formation (blue shading in Fig. 4a versus coloured dots in Fig. 3b,c); a reduction of sea ice concentration occurred both in the Alaskan and Siberian sectors in 2005 and 2007 (Fig. 4c and Extended Data Table 1). The September mean ice concentration dropped from 46% to 26% in the Alaskan sector and from 57% to 26% in the Siberian sector and has not recovered to date (Fig. 4c). Concurrently, the September mean sea surface temperature in these areas has risen from below 0 °C to 0.6 °C (Extended Data Fig. 1). These changes make it difficult for ice formed during a previous winter to survive the summer melt in these areas and survive into the following year. This is manifested in the drop of residence time of ice floes in areas of sea ice formation (Extended Data Fig. 2). The mean residence time of ice floes in the Siberian sector that arrived at the Fram Strait after 2007, was reduced from 15 to 6

months, indicating that most of the ice floes cannot survive the summer melt season and only new ice floes that formed after the summer enter the TPD.

**Fig. 4: Changes of sea ice concentration and sea ice motion.**




**a**,**b**, Difference of September sea ice concentration (**a**) and ice drift speed (**b**) between the two periods: 1990–2006 and 2007–2019. **c**,**d**, Time series of mean sea ice concentration in September (**c**) and mean sea ice drift speed in selected regions (**d**). **a**, The positive (negative) values indicate increase (decrease) in the latter period. **b**, The difference in sea ice drift vector is

shown by the arrows, while its magnitude is shown by the colour. The difference in sea ice drift field in **b** was calculated from ice drift vectors from December to May. The time series in **c** are the areal average of the Alaskan (A) and Siberian sectors (B) shown by the solid black polygon in **a**, while those in **d** are the areal average of A, B and C: the TPD Stream is shown by the rectangular box labelled C. The ice drift speed of the TPD in **d** shows the annual mean ice drift speed in box C (vector component parallel to the main axis of box C, positive value oriented to the Fram Strait), whereas those in A and B are calculated without three summer months (August to October) to exclude under-represented ice motion due to very low spatial coverage in recent years. **c**,**d**, The dashed lines indicate the detected regimes (Extended Data Table 1). Sea ice concentration from OSI SAF[51] and sea ice drift from Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors v.4.1 (ref. [52]) were used to derive the variables. The Matplotlib basemap toolkit was used to plot the map.

Acceleration of ice drift speed in the Arctic Ocean[27,28] has also contributed to the shorter residence time of sea ice. Figure 4b shows the difference in sea ice motion between the two periods, 1990–2006 and 2007–2019. Enhanced anticyclonic ice motion in the Canada Basin and acceleration of the TPD are clearly visible. These changes also exhibited a clear shift in 2007 (Fig. 4d and Extended Data Table 1). Westward ice drift speed in the Alaskan sector increased by 71% (from 2.0 cm $s^{-1}$ to 3.5 cm $s^{-1}$), northward ice drift in the Siberian sector increased by 42% (from 1.2 cm $s^{-1}$ to 1.7 cm $s^{-1}$) and the TPD accelerated by 37% (from 2.3 cm $s^{-1}$ to 3.2 cm $s^{-1}$), respectively. The acceleration of the TPD shortened the residence time of sea ice in the central Arctic Ocean by about four months on average. Concurrent shorter residence times in the ice formation areas and acceleration of the TPD led to a reduction of the mean residence time of sea ice by 1.6 years (from 4.3 to 2.7 years) after 2007.

## Shorter residence time and thinner ice

The relationship between the observed regime shift (Figs. 1b and 2) and reduced residence time (Fig. 3) is explained by a dynamic ice thickening process. Heat loss in open water areas to the atmosphere in autumn, followed by continued cooling in winter, forms a uniform ice thickness

distribution with a high modal peak[24] (thermodynamic forcing is governed by synoptic scales of $O(10^3)$ km). Dynamic forcing, on the other hand, increases the fraction of deformed thick ice forming the tail of the thickness distribution. Divergence, convergence and shear of wind and ocean currents cause mechanical fracturing of ice floes. When the forcing is convergent and once the local internal stress in the ice pack has exceeded the threshold, dynamic ice thickening occurs by ridging and/or rafting[25,29].

Dynamic ice thickening is a stochastic process because the deformation occurs in a small fraction of ice while the rest of the ice is unchanged when a dynamic event occurs. Ice thickness gained by an event (that is, ridging or rafting) also varies in space and differs between events. Another characteristic of dynamic thickening is its dependence on ice thickness. Thicker ice has a larger potential to get thicker as it can exert stronger compressive force on the ice forming ridges and rafting[30,31]. These characteristics enable us to formulate the process by a proportionate ice thickening of stochastic ice thickness, $X$, as,

$${X}_{i}={a}_{i-1}{X}_{i-1}$$
(1)

where $i$ denotes the time index counting dynamic ice thickening events and $a_{i-1}$ is the proportionate thickening increment due to the event at $i-1$. The increment represents the stochasticity of the dynamic thickening process. After $m$-times dynamic growth events, stochastic ice thickness is given by:

$${X}_{m}={X}_{0}{\Pi }_{i=0}^{i=m-1}{a}_{i}$$
(2)

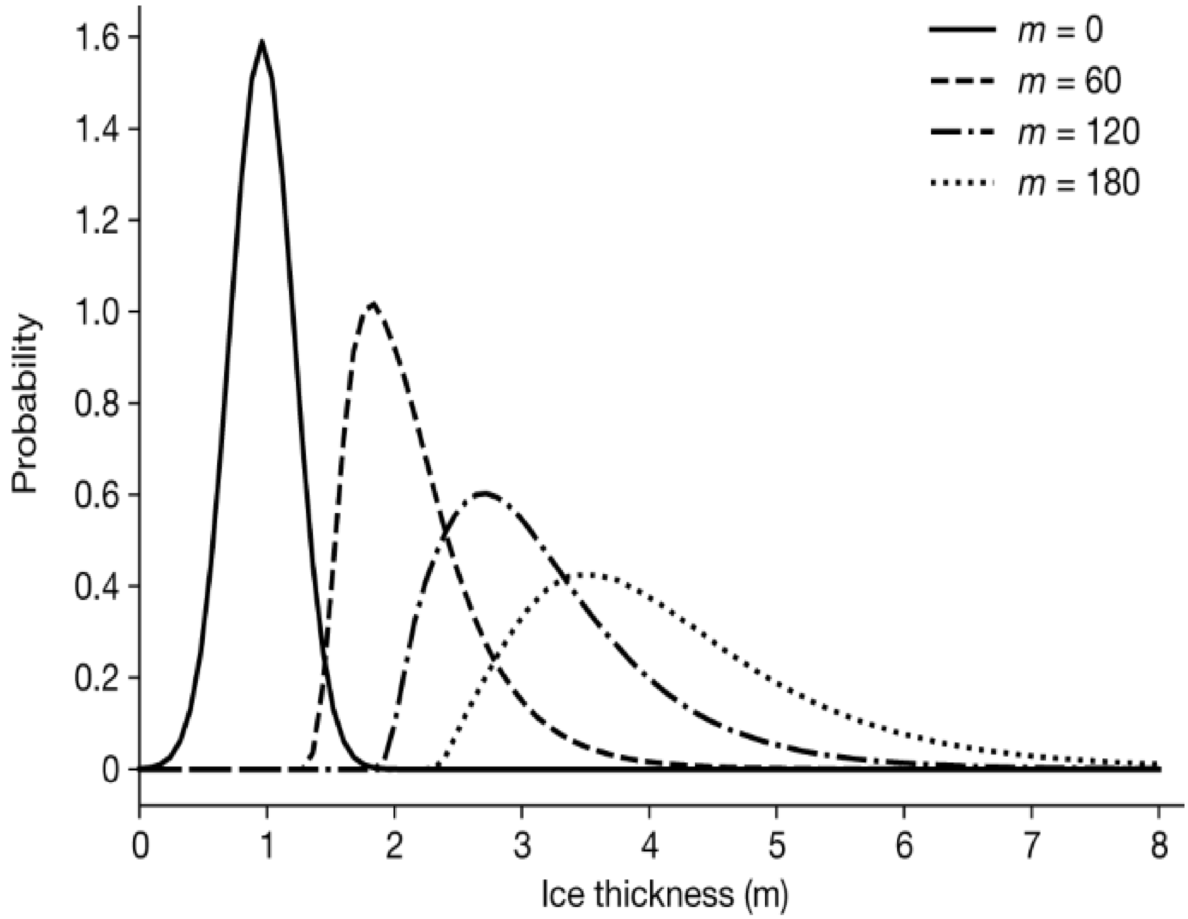
where $X_0$ is the initial ice thickness, that is, sea ice without dynamic thickening, and $\Pi$ is a product operator. Taking the natural logarithm of equation (2) gives:

$${\rm{ln}}\,{X}_{m}=\,{\rm{ln}}\,{X}_{0}+\,{\rm{ln}}\,{a}_{m-1}+\,{\rm{ln}}\,{a}_{m-2}+\ldots +\,{\rm{ln}}\,{a}_{0}$$
(3)

As $\ln X_0$, $\ln a_0$, $\ln a_1$, …, and $\ln a_i$ are uncorrelated (the thickening at each dynamic event is a stochastic process), the probability function of $\ln X_m$ gives a normal distribution for large $m$ (central limit theorem[32]), hence the probability function of $X_m$ gives a log-normal distribution (Gibrat's law[33]). The shape of the distribution varies with $m$, that is, the number of dynamic growth events. The larger the $m$, the larger the variance. If we assume that the number of dynamic growth events $m$ is proportional to the residence time of ice floes in the Arctic Ocean on annual timescales, a longer residence time leads to larger variance, that is, a larger fraction of deformed ice. As the number of such events has not considerably changed in the last three decades[34], $m$ can be linearly related to the residence time of sea ice in the Arctic Basin.

Figure 5 shows examples of the probability density function of $X_m$ for different $m$ (Methods). The plot demonstrates that the shorter residence (smaller $m$) gives higher modal peak and less deformed sea ice. The simple proportionate thickening process conceptually explains (1) the shape of the observed ice thickness distribution (the log-normal shapes in Fig. 1b) and, more importantly, (2) the reduction of the fraction of deformed ice for the shorter residence time after 2007 (Fig. 2).

**Fig. 5: Ice thickness distributions obtained from the stochastic model of dynamic ice thickening.**




Probability density functions of $X_m$ for different values of $m$. The smaller $m$ corresponds to a shorter residence time of sea ice in the Arctic Ocean. See the Methods for the descriptions.

## Discussion

The relationship between sea ice residence time in the Arctic Ocean and ice thickness distribution highlights the importance of coupled ocean–sea ice processes in the Alaskan and Siberian sectors of the Arctic (areas A and B in Fig. 3b). Several interrelated factors have become more prominent in the late twentieth century and have contributed to preconditioning the ocean–sea ice system before the stepwise changes in the ice formation areas: Arctic-wide rise of surface air temperature[35], thinning of sea ice[36], decrease of sea ice albedo[37] concurrent with a reduction of multi-year sea ice[38], increase of ocean heat flux through the Bering Strait[39] and increase of the upper ocean heat content[40]. September sea ice concentration in the Siberian sector

dropped below 40% in 2005 and the dramatic Arctic summer sea ice extent minimum occurred in 2007 (ref. [41]). This series of events initiated intensive and widespread ice–albedo feedback in the Alaskan and Siberian sectors in the summer[42,43], which resulted in a perennial increase of ocean heat content in areas of ice formation (Extended Data Fig. 1). After 2007, suppression of winter ice growth due to the accumulated ocean heat became conspicuous[44] and the resultant thinner ice pack became more vulnerable to summer melt in the following year. Prolongation of the summer melt season promoted further ice–albedo feedback[45] and has increased oceanic heat absorption in the summer[7]. Thus, summer ice extent and thickness in areas of ice formation has not recovered to the state before 2007 (Fig. 4c). In addition, continuing weakening of the cold halocline in the Siberian sector also influenced the upper ocean heat content[46] and possibly slowed down ice growth offshore of the Laptev Sea in recent years[17].

Our analysis demonstrates the long-lasting impact of climate change on Arctic sea ice through reduced residence time, suggesting an irreversible response of Arctic sea ice thickness connected to an increase of ocean heat content in areas of ice formation. The large reduction of summer ice extent in the Alaskan and Siberian sectors in 2005 and 2007 triggered intensive ice–albedo feedback[42,45] and initiated the perennial increase of ocean heat content in these areas[44]. This resulted in the stepwise reduction of residence time of sea ice in the Siberian sector of the Arctic, and hence a nonlinear response of the system. Before the shift, sea ice formed in and offshore of the Siberian shelves overwintered (spent about 15 months) in this area before entering the TPD (Extended Data Fig. 2), during which the ice thickened and increased its deformed fraction. After the shift, ice stayed in this area only about 6 months on average (Extended Data Fig. 2), resulting in recruitment of newly formed younger ice into the TPD and more sea ice formation during TPD transit to the Fram Strait[26]. The younger ice is thin, weakly linked and features ridges with more shallow keels; hence, it is more prone to wind forcing pushing the ice towards the Atlantic sector of the Arctic[28,47]. This process accelerated the TPD from 2007 onwards (Fig. 4d), while enhanced wind forcing after 2007 may also have contributed to the acceleration of the TPD (Extended Data Fig. 7). Because of the shorter residence time, the part of the ice that has thermodynamically grown is

thinner[17] (reduction of modal thickness in Fig. 1b) and the relative amount of the deformed fraction of ice has decreased (Figs. 1b and 2).

Impacts of this regime shift in Arctic sea ice on the pan-Arctic environment are extensive and require further investigation. Thinner and less deformed sea ice causes reduced momentum exchange between ice and ocean, contributing to reduced mixing in the upper ocean underneath areas that are fully covered with ice. This may affect entrainment of heat and nutrients from subsurface to surface ocean with a potential consequence on the biogeochemical cycles involving higher trophic levels. By contrast, however, sea ice retreat in marginal ice zones and continuing weakening of the cold halocline in the Atlantic sector allows for more turbulent mixing and winter convection in the upper ocean[46]. These counteracting effects can influence the regional contrasts of the ocean environment between fully ice-covered areas and marginal ice zones in the Arctic. In addition, habitat conditions of younger, level sea ice are different from those in older multi-year ice and might affect the sympagic (ice-associated) communities and their diversity[48,49]. Ridged sea ice supports higher biomass[48] and represents safe havens for organisms to hide from predators[50]. The amount of ridges and deformed ice has also consequences for human activity. Thinner, more level ice is less challenging for ship navigation than in thicker, deformed ice and, along with less ice/shorter ice seasons in general, may allow for an increase in Arctic maritime traffic. Finally, interdisciplinary studies in the Atlantic sector of the Arctic and downstream of the Fram Strait outflow are needed to shed light on the consequences of the described sea ice regime shift and its impacts on physical and biogeochemical processes.

# Methods

## Data

Sea ice draft data were obtained from upward looking sonar (ULS) moored in the East Greenland Current in the western Fram Strait. The dataset continuously covers the last three decades (1990–2019) with some short temporal gaps. Four ULSs were zonally aligned at approximately 79° N from 3° W to 6.5° W (Fig. 1a). The latitude of the mooring array changed from 79° N to 78.8° N in 2001. The zonal positions (names) of the moorings

equipped with ULS are 3° W (F11), 4° W (F12), 5° W (F13) and 6.5° W (F14), respectively. There are three main temporal data gaps during the three decades of measurements, that is, 1996, 2002 and 2008. Except for these gaps, ULSs were in operation although their number varied from time to time. The ULS measures the travel time of the sound reflected at the bottom of the floating sea ice, from which we calculate the ice draft, the underwater fraction of sea ice[54]. The raw data were processed to ice draft using procedures described in earlier literature[55,56]. The accuracy of each draft measurement ranges from 0.1 m (ice profiling sonars (IPS) deployed after 2006) to 0.2 m (ES300 instruments before 2006), while the uncertainty of each individual measurement is not subject to bias errors and the summary error statistics of monthly values are less than 0.1 m[57].

The daily mean sea ice motion product provided by the National Snow and Ice Data Center (Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors v.4.1, hereafter NSIDCv4 ice drift) was applied for backward trajectory calculation of sea ice floes, estimating residence time of sea ice in the Arctic and analysis of spatial average of ice drift speed. The product was derived from a combination of various ice motion estimates from remote sensing and tracks of ice-tethered buoys[52]. We applied the motion vectors from 1984 to 2019 for the backward trajectory calculation. Sea ice concentration data were taken from the Global Sea Ice Concentration Climate Data Records of the EUMETSAT OSI SAF[51] (OSI-409 v.1.2 for the period from January 1984 to April 2015 and OSI-430 v.1.2 for the period after April 2015). The ice concentration data were used for the sea ice trajectory calculations and to analyse spatial and temporal changes of ice concentration.

Sea surface temperatures obtained from the NOAA/NESDIS/NCEI Daily Optimum Interpolation Sea Surface Temperature v.2.1 (ref. [58]) were used to examine temporal variation in sea surface temperature in areas of ice formation (Extended Data Fig. 1a,b). A dataset of Arctic Ocean in situ hydrographic observations[59] was also used to examine the change in ocean surface temperature between 1990 and 2006 and between 2007 and 2019. The dataset was revised using recent observations[60] and gridded to 110 × 110 km cells covering the whole Arctic Ocean. The mean temperature in each period (1990–2006 and 2007–2019) in each cell was defined by an

average of all available measurements for every 3-month period (January–March, April–June, July–September and October–December). If the number of available data in a cell was less than four, a missing value was assigned to the cell. The difference in summer sea surface temperature (July–September, 0–20 m) between two periods (1990–2006 and 2007–2019) was used for the analysis (Extended Data Fig. 1c). Atmospheric data were taken from the European Center for Medium-Range Weather Forecasts Reanalysis v.5 (ref. [61]) (ERA5). Daily and monthly mean sea-level pressure and 10 m wind were used in the analyses.

## Ice thickness distribution

Ice thickness distributions were derived on a monthly basis. All sea ice draft measurements in the Fram Strait from 1990 to 2019 were classified into draft thickness bins of 0.1 m, ranging from 0 to 8 m (80 bins in total). The number of data samples (ice draft measurements) used to derive the distributions varied from time to time. From 1990 to 2005, $O(10^4)$ samples were used to derive the distribution functions on a monthly basis (measurement interval of 240 s in most cases); after 2006, $O(10^6)$ samples were used (interval of 2 s). The number of samples used were sufficient to derive thickness distribution on a monthly basis[57]. The number of data in each bin were divided by the total number of measurements to derive the distribution function. The open water fraction (that is, zero ice thickness) was excluded when deriving the function. Distribution functions, including the open water fraction, are shown in Extended Data Fig. 3. If the temporal coverage of the data samples was less than 15% of a monthly coverage, the distribution function was not defined and removed from the analysis. The draft thickness distributions were converted to ice thickness distributions by an average ratio of draft to thickness in the Fram Strait, 1.136 (ref. [62]). Although the ratio has some seasonal variability and might have slightly changed due to changes in ice density and snow load in different seasons and years, we assumed that the change was not considerable for the aim of the current analyses. A composite time series of ice thickness distribution (Fig. 2a and Extended Data Fig. 3) was obtained from a combination of available distribution functions from F11 to F14 . More specifically, one of the distribution functions from F11 to F14 was used with a priority order of F13, F14, F12, F11. The composite time series is shown in Fig. 2a, while the time

series at each site are shown in Extended Data Fig. 4. The fractions of sea ice thicker than two thresholds, 4 and 5 m, were calculated as the cumulative function of all available distributions (that is, all distributions from F11 to F14) in each month and are shown in Fig. 2d.

The uncertainties on the estimates of monthly fractions of ice thicker than a threshold and position of the modal ice peak were assessed numerically using a moving block bootstrap approach[63]. Bootstrapping is a family of resampling techniques used to derive uncertainties on various complex estimators for large datasets and uses random sampling with replacement[64]. The presence of autocorrelation in ice draft series from IPS (ULSs deployed after 2006) suggested using the moving block bootstrap approach[63]. The method splits the original monthly series of ice drafts $O(10^6)$ samples of length $N$ into $N - K + 1$ overlapping blocks of length $K$ each. The block length was set to 30 samples that approximately corresponded to a distance of 10 m covered by ice travelling at 0.3 knots. It roughly corresponded to the lower limit of a horizontal spatial scale of ice ridges. For ES300 instruments (ULSs deployed until 2005) with a lower sampling rate of 240 s, ordinary bootstrapping was used.

At each of $M$ steps of bootstrap sampling, $N/K$ blocks were drawn at random, with replacement, from the constructed set of $N - K + 1$ blocks, making a new bootstrap data sample for the month. A Gaussian noise of $N(0,1)$ was further added to the data to account for measurement uncertainty. The mean and s.d. of the fractions of thicker ice and modal ice peak position were then calculated directly from the $M$ estimates derived at each step of the procedure. The results suggested that the monthly coefficient of variation or a ratio of the s.d. of the estimate to its mean, for the IPS data varies from 1% to 3% for both fractions, being on average lower (1–2%) for a fraction of ice greater than 4 m thick. For ES300, the coefficient of variation was slightly higher at about 4(6)% for the fractions of ice thicker than 4(5) m. The same applied to the position of the modal peak, which showed a coefficient of variation of 0–3% being typically closer to 0 for the IPS data. It suggested that the selected bin width was large enough to accommodate uncertainties related to the approach and data. For the ES300 data, the monthly s.d. of the modal peak position was higher, up to 30 cm, and the coefficient of variation was within 9%. Therefore, we postulate that the

inferred uncertainties are far too low to have any noticeable influence on the results of the shift detection analysis.

## Modal peak and variance of ice thickness distributions

As statistical measures of ice thickness distributions, we examined modal thickness and variance. The monthly mean ice thickness distributions (740 samples in total) were fitted to log-normal functions:

$$F(x)=\frac{1}{\sqrt{2{\rm{\pi }}{\sigma }_{{\rm{f}}}^{2}}x}\exp \left(\frac{-{({\rm{l}}{\rm{n}}x-{\mu }_{{\rm{f}}})}^{2}}{2{\sigma }_{{\rm{f}}}^{2}}\right)$$
(4)

where $\sigma_{\rm f}$ and $\mu_{\rm f}$ are the fitting parameters, $x$ is the ice thickness bin and $F$ is the distribution function. To detect the second peak of the distributions that represents multi-year ice travelled across the Arctic Basin, a cut-off threshold was introduced. The threshold was used to exclude thin sea ice fraction, which is supposedly formed in the vicinity of the Fram Strait and is not representing basin-wide changes of ice properties in the Arctic. We defined the threshold by the minimum between the first and second peak of each monthly mean distribution. A set of two consecutive negative gradients (towards thicker bins) followed by two consecutive positive gradients was used to detect the minimum (after applying 3-bin smoothing), while a threshold of 1.53 m (corresponding to 1.3 m of ice draft) was applied when an estimated threshold was thicker than 3 m. Function values ranging lower than the threshold were set to zero (zero case) or excluded from the fitting (NaN case). A least-square minimization was applied to fit a log-normal function to the distribution. In general, the fitted log-normal functions represent the distribution very well. The NaN case slightly underestimates the modal peak, while the zero case captures the peak very well. Examples of distribution functions, together with cut-off thresholds and fitted log-normal functions, are shown in Extended Data Fig. 5. The modal peak of the log-normal function roughly gives the thickness of thermodynamically grown sea ice, while the variance of the function quantifies the deformed fraction of sea ice (dynamically thickened thickness). Changes in modal peak height and variance of the fitted log-normal functions, var($x$) = exp

$(2\mu_f + \sigma_f^2)$ $(\exp(\sigma_f^2) - 1)$, for the last three decades are summarized in Fig. 2b,c. The time series of modal thickness and the fitting parameters $\sigma_f$ and $\mu_f$ are summarized in Extended Data Fig. 6.

## Regime shifts detection

A sequential algorithm for regime shift detection[23] was applied to all time series. The method identifies discontinuities in a time series using a data-driven approach that does not require an a priori assumption on the timing of the regime shifts. The method first identifies potential change points sequentially by checking if the anomaly of the data point is statistically significant from the mean value of the current regime. If it is significant, the following data points are sequentially used to assess the confidence of the shift, using a regime shift index (RSI). RSI represents a cumulative sum of normalized deviations from the hypothetical mean level for the new regime, for which the difference from the mean level of the current regime is statistically significant according to a Student's *t*-test. If the RSI is positive for all points sequentially within the specified cut-off length, the null hypothesis of a constant mean is rejected. This led us to conclude that the regime shift might have occurred at that point in time[65]. If multiple data are available at a certain point in time (that is, multiple sites from F11 to F14), the mean value is applied in the time series. Before testing, the temporal gaps of the time series were interpolated by the average of all available data (modal peak height (Fig. 2b), variance (Fig. 2c) and modal thickness (Extended Data Fig. 6a) of ice thickness distributions, fraction of thick sea ice (Fig. 2d) and residence time of sea ice (Fig. 3a and Extended Data Fig. 2)). The cut-off length was set to 7 years (84 months) to cover the advection timescale (travelling time across the Arctic) of sea ice, while at the same time, detecting shifts occurring at a timescale shorter than a decade. Other cut-off lengths (3, 4, 5, 6, 8 and 10 years) were also tested to see the sensitivity and robustness of the results. A summary of the test results is given in Extended Data Table 1; the timing of the detected shifts is shown in all time series except for Fig. 2. The timing of the detected shifts of modal peak height and variance of ice thickness distributions are shown in Extended Data Fig. 6b,c, while those of the fraction of thick sea ice are shown in Extended Data Fig. 8. RSIs, respective *P* values and the shift of the means are summarized in Extended Data Table 2.

## Sea ice trajectory analysis

To investigate changes of pathways and residence time of sea ice in the Arctic basins, sea ice trajectories were calculated for the last three decades. Eight pseudo-ice floes were settled in the western half of the Fram Strait section (from prime meridian to 10° W) at the same time and advected backwards in time. The calculations started on the 15th of every month from 1990 to 2019. Daily sea ice motion vectors from the NSIDCv4 were used to update daily position of ice floes backwards in time. Ice motion vectors at the respective floe positions were calculated by interpolation of surrounding points with Gaussian-type weighting (e-folding scale of 25 km). Each trajectory calculation was performed 6 years back in time, while it was terminated if no motion vector was available within a 25-km distance or sea ice concentration at the floe position was lower than 15%. The sea ice concentration at the ice floe position was obtained from OSI-409/OSI-430 with a Gaussian-type weighting (e-folding scale of 12.5 km). The position of each trajectory termination was used to define the location of 'initial sea ice formation'. Trajectories shorter than three months were excluded from the analysis because they represent ice floes formed in the vicinity of the Fram Strait.

Uncertainty of the daily position of the pseudo-floes was assessed by comparisons with ice-tethered buoy tracks obtained from the International Arctic Buoy Program (IABP)[53]. We used 83 buoy tracks that arrived in the Fram Strait from 2000 to 2018 and calculated the corresponding pseudo-buoy tracks backwards in time. The comparisons showed that the mean error of the daily pseudo-buoy positions can be reasonably approximated by a linear function of backtracking days[19], error = 50 + (backtracking days)/2 km. We applied this empirical formula as an error of the daily position of the backward trajectories from 0 to 500 backtracking days, which corresponds to a 200 (300) km error after 300 (500) backtracking days. Note that this error estimate may underestimate the uncertainty because IABP buoy tracks have been included in the NSIDCv4 ice motion product. However, comparisons between non-IABP buoys and pseudo-buoy tracks derived from the NSIDCv4 with error estimates by a bootstrap method showed that pseudo-tracks are largely parallel to the corresponding buoys and the error does not monotonically increase over time[66]. The estimated error circles

(approximately 300 km) of ice formation location in the present study are sufficiently small compared to the polygons in Fig. 3b (greater than 1,500-km width), which guarantees the robustness of the analysis.

The residence time of sea ice in the Arctic basins was defined by the period from the start to the termination date of each trajectory. We calculated an average of residence time of eight pseudo-ice floes that arrived in the Fram Strait at the same time and used it to define the mean residence time of ice floes for each month. The uncertainty of the residence time was defined by the s.d. of the residence time of the eight pseudo-ice floes.

## Stochastic model of dynamic ice thickening

The log-normal form of the ice thickness distribution can be obtained from a simple proportionate growth process, ${X}_{m}={X}_{0}{\Pi }_{i=0}^{i=m-1}{a}_{i}$. If ln $X_0$ and ln $a_0$, ln $a_1$, .., ln $a_i$ are uncorrelated; thus, the probability function of $X_m$ for large $m$ is given by:

$$f({X}_{m})=\frac{1}{\sqrt{2{\rm{\pi }}}\mathop{\sigma }\limits^{ \acute{} }{(m)}^{2}}{X}_{m}}\exp \left(\frac{-{({\rm{l}}{\rm{n}}({X}_{m}/\mathop{X}\limits^{ \acute{} }(m)))}^{2}}{2\mathop{\sigma }\limits^{ \acute{} }{(m)}^{2}}\right)$$
(5)

where $\acute{X}(m)={e}^{\nu m}$ and $\acute{\sigma }(m)=\sigma {m}^{1/2}$ and $v$ and $\sigma^2$ are the mean and variance of the population distribution of ln $a_m$ (including $X_0$), respectively[67].

In this study, we provide a concept and description of a stochastic model that formulates sea ice thickening associated with dynamic ice deformation. The model formulates three features of dynamic sea ice thickening by ridging and/or rafting: (1) dynamic ice thickening is a stochastic process (areal and thickening stochasticity); (2) thicker ice has a larger potential to get thicker than thinner ice at a dynamic event (proportionate ice thickening); and (3) thinner ice has a higher probability of dynamic deformation due to its weaker ice strength (preferential deformation of thinner over thicker ice types). The first point consists of two types of stochasticity in the dynamic

ice thickening process. One is areal stochasticity, corresponding to the fact that ice deformation only occurs for a small fraction of the pack ice while the rest of the ice is unchanged when a dynamic event occurs. The other is thickening stochasticity, representing the fact that the thickness gain by ridging/rafting varies in space and differs between events. The second point represents a sea ice characteristic that thicker ice is tolerant and can exert stronger compressive force on the ice forming ridges and/or rafts; hence, more energy is potentially available for the dynamic thickening[30,31]. The third point represents the fact that the thinner part of the pack ice is preferentially ridged/rafted when a dynamic event occurs[68,69]. This also takes into account the effect of ice thickness changes on the dynamic thickening process, for example, the thinner ice condition in recent years increases the likelihood of ice deformation.

The two first points can be formulated by a proportionate thickening of stochastic ice thickness:

$${X}_{i}={a}_{i-1}{X}_{i-1}$$
(6)

where $X$ is the stochastic ice thickness at a certain location, $i$ denotes the time index counting sporadic dynamic events (for example, passage of a storm) and $a_{i-1}$ is the conditional proportionate thickening increment due to the event at $i-1$. The increment $a_{i-1}$ is a stochastic variable representing both areal and thickening stochasticity. They are implemented as:

$${a}_{i}=\left\{\begin{array}{cc}1+b{r}_{i} & {\rm{w}}{\rm{i}}{\rm{t}}{\rm{h}}\,{\alpha }_{i}({\rm{ \% }})\,{\rm{p}}{\rm{r}}{\rm{o}}{\rm{b}}{\rm{a}}{\rm{b}}{\rm{i}}{\rm{l}}{\rm{i}}{\rm{t}}{\rm{y}}\\ 1 & {\rm{w}}{\rm{i}}{\rm{t}}{\rm{h}}\,1-{\alpha }_{i}({\rm{ \% }})\,{\rm{p}}{\rm{r}}{\rm{o}}{\rm{b}}{\rm{a}}{\rm{b}}{\rm{i}}{\rm{l}}{\rm{i}}{\rm{t}}{\rm{y}}\end{array}\right.$$
(7)

where $b$ is a proportionate thickening constant, $r_i$ is a stochastic thickening increment that represents the thickening stochasticity of $i$-th dynamic event and $\alpha_i$ is the areal probability of dynamic thickening that represents areal

stochasticity and gives the probability of ice thickening occurrence. This formula indicates that when a dynamic event occurs, $\alpha(\%)$ area of pack ice experiences dynamic thickening (ridging/rafting), while the rest $(1 - \alpha(\%))$ is unchanged (areal stochasticity). The thickness gain, $br_i$, in the dynamic thickening area, is also a stochastic variable: the possible maximum gain is $b$ while the minimum is 0 ($r_i$ is random, so $0 \le r_i < 1$) (thickening stochasticity).

We applied the proportionate thickening constant as $b = 0.4$. The choice of $b = 0.4$ implies that sea ice in the ridging/rafting area gains 0.4 m thickness at maximum (0.2 m on average) when the ice is initially 1-m thick, while it gains a 1.2 m thickness at maximum (0.6 m on average) when initially the ice is 3-m thick. The value of the parameter $b$ comes from a recent high-resolution observation (5 × 5 m resolution covering a 9-km$^2$ area) of single ice deformation event north of Svalbard, which describes changes in the sea ice freeboard just before and right after a storm event[70]. According to this study, the change in sea ice freeboard in a converging area is 0.07 m (from 0.36 m to 0.43 m) on average, corresponding to 0.58 m dynamic thickening of ice by ridging and/or rafting (assuming the freeboard to thickness ratio = 8.35)[62]. The gain relative to the mean ice thickness is estimated by $b = 0.58/1.45 = 0.4$ (the mean ice thickness in the survey area = 1.45 m). We applied the value for the proportionate thickening constant, $b$, as the first approach to develop the model. Although the study captured detailed spatial change in the sea ice freeboard, the estimate of $b$ comes from a one-time event and hence needs further assessments by future observations. It should be noted that $b = 0.4$ is an areal average estimated from the observations, whereas we applied $b$ as the upper bound of the proportionate thickening. This is because the probability function of the stochastic thickening increment, $r_i$, is not known so far; hence, we assumed a constant probability of the increment between 0 and $b$, potentially causing excessive thickening near the upper bound. The effect of the choice of $b$ is discussed below.

The areal probability of dynamic thickening, $\alpha$, is included to take the third point into account, that is, thinner ice has more chance to be ridged and/or rafted than thicker ice when a dynamic event occurs. To implement this feature, $\alpha$ is given by a function of ice thickness: the areal probability is inversely proportional to the stochastic ice thickness $X_i$:

$${\alpha }_{i}({X}_{i})=\frac{8}{{X}_{i}+1}({\rm{ \% }}).$$
(8)

The formula indicates that 1-m thick ice experiences dynamic thickening at 4% areal probability, while 3-m ice experiences dynamic thickening at 2% areal probability. Our first implementation of this formula is based on an observational estimate of areal fraction of dynamic thickening[70]. According to the high-resolution survey of a single dynamic event, thickening occurred in 4% of the survey area with a mean ice thickness of 1.45 m. This formula also needs further evaluation by comparing with future observations that address the relationship between areal probability of deformation and ice thickness.

Another parameter necessary for the model is the number of dynamic events, *m*, that is, external forcing that could cause mechanical fracturing of sea ice and consequent ridging and/or rafting. We used the number of Arctic cyclones that passes over the ice pack as a first-order indicator of the number of dynamic events. Typically 90–130 cyclones per year occur in the Arctic Ocean (40–60 cyclones in winter, 50–70 cyclones in summer)[71]. A typical size of an Arctic cyclone is approximately $3 \times 10^6$ km$^2$ (mean radius of approximately $10^3$ km)[71], which covers approximately one third of the ice-covered area of the Arctic Ocean. We therefore assumed that one-third of all cyclones hits the ice pack at a certain location in the Arctic, that is, the ice pack experiences approximately 40 dynamic events per year. This corresponds to approximately 80–240 dynamic deformation events for the typical residence time of sea ice in the Arctic (2–6 years; Fig. 3a).

In addition, examples shown in Fig. 5 contain a simple thermodynamic term to mimic the effect of modal peak shift of thickness distribution due to thermodynamic ice growth:

$${X}_{i}={a}_{i-1}{X}_{i-1}+c/{X}_{i-1}$$
(9)

where *c* is the thermodynamic ice growth coefficient. This term comes from a simplified thermodynamic process without thermal inertia of sea ice and heat flux from the ocean[72]:

$$\frac{{\rm{d}}H}{{\rm{d}}t}=\frac{{\kappa }_{{\rm{ice}}}({T}_{{\rm{f}}}-{T}_{{\rm{s}}})}{{\rho }_{{\rm{ice}}}{L}_{{\rm{f}}}H}$$
(10)

where $H$ is the ice thickness, $\kappa_{\rm ice}$ is the heat conductivity of ice, $\rho_{\rm ice}$ is the ice density, $L_{\rm f}$ is the latent heat of freezing, $T_{\rm f}$ is the freezing temperature of sea ice and $T_{\rm s}$ is the temperature at the ice surface. We applied this formula with a simplification, $\Delta H = c/H$, where $\Delta H$ is the ice thickness change due to a thermodynamic process, $c$ is a thermodynamic ice growth coefficient corresponding to $\Delta t\kappa_{\rm ice}\,(T_{\rm f} - T_{\rm s})/(\rho_{\rm ice}L_{\rm f})$. As the model does not include a process that forms new thin ice by lead opening, inclusion of the thermal forcing term without a compensating term makes the modal peak very steep after few years, that is, no ice exists in thickness ranges thinner than the thermal equilibrium thickness. To alleviate such an excessive modal peak generation and to take into account the insulating effect of the snow pack that substantially delays thermodynamic ice growth, we applied a moderate value, $c = 0.015$, which is about one-third of the value estimated from $c = \Delta t\kappa_{\rm ice}\,(T_{\rm f} - T_{\rm s})/(\rho_{\rm ice}L_{\rm f})$, where $\Delta t \cong 9$ d (corresponding to 40 dynamic events per year) and annual mean surface air temperature of $T_{\rm s} = 263$ K.

The initial condition of the ice thickness distribution in Fig. 5 is given by a thermodynamically grown sea ice without dynamic deformation, $X_0$. This is also a stochastic variable, having a normal distribution for simplicity:

$$g({X}_{0})=\frac{1}{\sqrt{2{\rm{\pi }}}{\sigma }_{0}}\exp \left(\frac{-{(x-{\mu }_{0})}^{2}}{2{\sigma }_{0}^{2}}\right)$$
(11)

where $\mu_0 = 1.0$ and $\sigma_0 = 0.25$ are applied in the examples (that is, 1 m mean ice thickness with 0.25 s.d., shown by $m = 0$ in Fig. 5), which roughly corresponds to the thickness of new ice three months after its formation (based on Anderson's freezing degree days law[25], with an assumption of $T_s = 253$ K). Figure 5 shows examples of ice thickness distribution after 60, 120 and 180 dynamic events, roughly corresponding to 1.5, 3 and 4.5 years of residence time of sea ice.

The current formulation contains three parameters to describe the dynamic ice thickening process: $b$ (the proportionate thickening constant); $\alpha$ (the areal probability of dynamic thickening); and $m$ (the number of dynamic events). In this study, we briefly describe the sensitivity of the ice thickness distributions to these parameters. In general, a smaller (larger) thickening constant $b$ decelerates (accelerates) the dynamic thickening process, that is, a smaller $b$ gives a smaller variance and steeper modal peak of thickness distribution if $\alpha$ and $m$ are fixed. However, a large value of $b$ (for example, $b = 0.8$, indicating that ridged ice can be 1.8 times thicker than the ice before an event at maximum) makes the distribution bimodal because the possible thickness gain at each dynamic event is far from the modal thickness and the ridged/rafted ice tends to generate another peak apart from the mode. Therefore, the possible and realistic range of $b$ should be examined further together with the probability density function of the thickening increment $r$ by high-resolution observations in the future. The areal probability of dynamic thickening, $\alpha$, also affects the evolution of the dynamic thickening process. A larger $\alpha$ promotes dynamic thickening because a larger fraction of pack ice can be deformed at one event. The thickness dependency of the probability, equation (8), decelerates further thickening of thick ice. Although values of $b$ and $\alpha$ affect the progress of dynamic thickening in the model, we obtained similar ice thickness distributions with a log-normal form sooner or later, that is, smaller $b$ and $\alpha$ can be compensated by a large $m$, the number of deformation events, indicating a robustness of the formulation. The resulting shape of the distribution, its temporal evolution (Fig. 5) and its comparison with the observed change in distribution (Fig. 1b) suggest that the proposed stochastic ice thickening model captures the essence of the dynamic thickening process that resulted in the observed changes in ice thickness distribution.

# Data availability

The ice thickness distribution data from the Fram Strait Arctic Outflow Observatory are available from the Norwegian Polar Data Centre, https://data.npolar.no/dataset/b94cb848-3120-4f29-a827-298108e0d059. The ice drift data are available at the National Snow and Ice Data Center https://nsidc.org/data/NSIDC-0116/versions/4 (NSIDCv4). The sea ice concentration data used are available at

ftp://osisaf.met.no/reprocessed/ice/conc/v2p0/ (OSI-409, superseded by OSI-450) and ftp://osisaf.met.no/reprocessed/ice/conc-cont-reproc/v1p2/ (OSI-430). The ERA5 reanalysis product is available at https://cds.climate.copernicus.eu/#!/home.

## Code availability

The backward trajectory code is available at *Zenodo* https://zenodo.org/record/7390660#.Y4oSL9LMIUF (https://doi.org/10.5281/zenodo.7390659).

## References


1. Fox-Kemper, B. et al. Ocean, cryosphere and sea level change. *IPCC Climate Change 2021: The Physical Science Basis* (eds Masson-Delmotte, V. et al.) (Cambridge Univ. Press, 2021).

2. Cooper, G. S., Willcock, S. & Dearing, J. A. Regime shifts occur disproportionately faster in larger ecosystems. *Nat. Commun.* **11**, 1175 (2020).

3. Druckenmiller, M. L., Moon, T. & Thoman, R. (eds) *State of the Climate in 2020: The Arctic*. Bull. Am. Meteorol. Soc. 102, S263–S315 (American Meteorological Society, 2020).

4. Frey, K. E., Moore, G. W. K., Cooper, L. W. & Grebmeier, J. M. Divergent patterns of recent sea ice cover across the Bering, Chukchi, and Beaufort Seas of the Pacific Arctic Region. *Prog. Oceanogr.* **136**, 32–49 (2015).

5. Serreze, M. C., Holland, M. M. & Stroeve, J. Perspectives on the Arctic's shrinking sea-ice cover. *Science* **315**, 1533–1536 (2007).

6. Perovich, D. K. et al. Increasing solar heating of the Arctic Ocean and adjacent seas, 1979–2005: attribution and role in the ice-albedo feedback. *Geophys. Res. Lett.* **34**, L19505 (2007).

7. Stroeve, J. C., Markus, T., Boisvert, L., Miller, J. & Barrett, A. Changes in Arctic melt season and implications for sea ice loss. *Geophys. Res. Lett.* **41**, 1216–1225 (2014).

8. Colony, R. & Thorndike, A. S. An estimate of the mean field of Arctic sea ice motion. *J. Geophys. Res.* **89**, 10623–10629 (1984).

9. Pfirman, S. L., Colony, R., Nürnberg, D., Eicken, H. & Rigor, I. Reconstructing the origin and trajectory of drifting Arctic sea ice. *J. Geophys. Res.* **102**, 12575–12586 (1997).

10. Serreze, M. C., Barry, R. G. & McLaren, A. S. Seasonal variations in sea ice motion and effects on sea ice concentration in the Canada Basin. *J. Geophys. Res.* **94**, 10955–10970 (1980).

11. Ricker, R. et al. Evidence for an Increasing role of ocean heat in Arctic winter sea ice growth. *J. Clim.* **34**, 5215–5227 (2021).

12. Martin, T. & Augstein, E. Large-scale drift of Arctic Sea ice retrieved from passive microwave satellite data. *J. Geophys. Res.* **105**, 8775–8788 (2000).

13. Maslanik, J., Stroeve, J., Fowler, C. & Emery, W. Distribution and trends in Arctic sea ice age through spring 2011. *Geophys. Res. Lett.* **38**, L13502 (2011).

14. Krumpen, T. et al. Recent summer sea ice thickness surveys in Fram Strait and associated ice volume fluxes. *Cryosphere* **10**, 523–534 (2016).

15. Serreze, M. C. et al. The large-scale freshwater cycle of the Arctic. *J. Geophys. Res.* **111**, C11010 (2006).

16. Haine, T. W. N. et al. Arctic freshwater export: status, mechanisms, and prospects. *Glob. Planet. Change* **125**, 13–35 (2015).

17. Belter, J. H. et al. Interannual variability in Transpolar Drift summer sea ice thickness and potential impact of Atlantification. *Cryosphere* **15**, 2575–2591 (2021).

18. de Steur, L., Peralta-Ferriz, C. & Pavlova, O. Freshwater export in the East Greenland Current freshens the North Atlantic. *Geophys. Res. Lett.* **45**, 13359–13366 (2018).

19. Sumata, H., de Steur, L., Gerland, S., Divine, D. V. & Pavlova, O. Unprecedented decline of Arctic sea ice outflow in 2018. *Nat. Commun.* **13**, 1747 (2022).

20. Vinje, T., Nordlund, N. & Kvambekk, Å. Monitoring ice thickness in Fram Strait. *J. Geophys. Res.* **103**, 10437–10449 (1998).

21. Hansen, E. et al. Variability in categories of Arctic sea ice in Fram Strait. *J. Geophys. Res. Oceans* **119**, 7175–7189 (2014).

22. Renner, A. H. H. et al. Evidence of Arctic sea ice thinning from direct observations. *Geophys. Res. Lett.* **41**, 5029–5036 (2014).

23. Rodionov, S. N. A sequential algorithm for testing climate regime shifts.*Geophys. Res. Lett.* **31**, L09204 (2004).

24. Maykut, G. A. & Untersteiner, N. Some results from a time-dependent thermodynamic model of sea ice. *J. Geophys. Res.* **76**, 1550–1575 (1971).

25. Wadhams, P. *Ice in the Ocean* (Gordon & Breach Science, 2002).

26. Krumpen, T. et al. Arctic warming interrupts the Transpolar Drift and affects long-range transport of sea ice and ice-rafted matter. *Sci. Rep.* **9**, 5459 (2019).

27. Spreen, G., Kwok, R. & Menemenlis, D. Trends in Arctic sea ice drift and role of wind forcing: 1992–2009. *Geophys. Res. Lett.* **38**, L19501 (2011).

28. Kwok, R., Spreen, G. & Pang, S. Arctic sea ice circulation and drift speed: decadal trends and ocean currents. *J. Geophys. Res.* **118**, 2408–2425 (2013).

29. Leppäranta, M. *The Drift of Sea Ice* 2nd edn (Springer-Verlag, 2011).

30. Melling, H. & Ridel, D. A. Development of seasonal pack ice in the Beaufort Sea during the winter of 1991–1992: a view from below. *J. Geophys. Res.* **101**, 11975–11991 (1996).

31. Hopkins, M. A. Four stages of pressure ridging. *J. Geophys. Res.* **103**, 21883–21891 (1998).

32. Lindeberg, J. W. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung [A new derivation of the exponential law in the probability calculation]. *Math. Z.* **15**, 211–225 (1922).

33. Crow, L. C. & Shimizu, K. (eds) *Lognormal distributions: Theory and Applications* (CRC Press, 2020).

34. Rinke, A. et al. Extreme cyclone events in the Arctic: wintertime variability and trends. *Environ. Res. Lett.* **12**, 094006 (2017).

35. Rantanen, M. et al. The Arctic has warmed nearly four times faster than the globe since 1979. *Commun. Earth Environ.* **3**, 168 (2022).

36. Lindsay, R. & Schweiger, A. Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. *Cryosphere* **9**, 269–283 (2015).

37. Perovich, D. K. & Polashenski, C. Albedo evolution of seasonal Arctic sea ice. *Geophys. Res. Lett.* **39**, L08501 (2012).

38. Maslanik, J. A. et al. A younger, thinner Arctic ice cover: increased potential for rapid, extensive sea-ice loss. *Geophys. Res. Lett.* **34**, L24501 (2007).

39. Woodgate, R. A., Weingartner, T. & Lindsay, R. The 2007 Bering Strait oceanic heat flux and anomalous Arctic sea-ice retreat. *Geophys. Res. Lett.* **37**, L01602 (2010).

40. Steele, M., Ermold, W. & Zhang, J. Arctic Ocean surface warming trends over the past 100 years. *Geophys. Res. Lett.* **35**, L02614 (2008).

41. Stroeve, J. et al. Arctic sea ice extent plummets in 2007. *Eos* **89**, 13–14 (2008).

42. Perovich, D. K., Richter-Menge, J. A., Jones, K. F. & Light, B. Sunlight, water, and ice: extreme Arctic sea ice melt during the summer of 2007. *Geophys. Res. Lett.* **35**, L11501 (2008).

43. Lei, R. et al. Changes in summer sea ice, albedo, and portioning of surface solar radiation in the Pacific sector of Arctic Ocean during 1982–2009. *J. Geophys. Res. Oceans* **121**, 5470–5486 (2016).

44. Timmermans, M.-L. The impact of stored solar heat on Arctic sea ice growth. *Geophys. Res. Lett.* **42**, 6399–6406 (2015).

45. Kashiwase, H., Ohshima, K. I., Nihashi, S. & Eicken, H. Evidence for ice-ocean albedo feedback in the Arctic Ocean shifting to a seasonal ice zone. *Sci. Rep.* **7**, 8170 (2017).

46. Polyakov, I. V. et al. Weakening of cold halocline layer exposes sea ice to oceanic heat in the eastern Arctic Ocean. *J. Clim.* **33**, 8107–8123 (2020).

47. Olason, E. & Notz, D. Drivers of variability in Arctic sea-ice drift speed. *J. Geophys. Res. Oceans* **119**, 5755–5775 (2014).

48. Hop, H. et al. Ice amphipod distribution relative to ice density and under-ice topography in the northern Barents Sea. *Polar Biol.* **23**, 357–367 (2000).

49. Olsen, L. M. et al. The seeding of ice algal blooms in Arctic pack ice: the multiyear ice seed repository hypothesis. *J. Geophys. Res. Biogeosci.* **122**, 1529–1548 (2017).

50. Gradinger, R., Bluhm, B. & Iken, K. Arctic sea-ice ridges—Safe heavens for sea-ice fauna during periods of extreme ice melt? *Deep Sea Res. 2 Top. Stud. Oceanogr.* **57**, 86–95 (2010).

51. Lavergne, T. et al. Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records. *Cryosphere* **13**, 49–78

(2019).

52. Tschudi, M. A., Meier, W. N. & Stewart, J. S. An enhancement to sea ice motion and age products at the National Snow and Ice Data Center (NSIDC). *Cryosphere* **14**, 1519–1536 (2020).

53. Rigor, I. G., Wallace, J. M. & Colony, R. L. Response of sea ice to the Arctic Oscillation. *J. Clim.* **15**, 2648–2663 (2002).

54. Melling, H., Johnston, P. H. & Riedel, D. A. Measurements of the underside topography of sea ice by moored subsea sonar. *J. Atmos. Ocean. Technol.* **12**, 589–602 (1995).

55. Fissel, D. B., Marko, J. R. & Melling, H. Advances in upward looking sonar technology for studying the processes of change in Arctic Ocean ice climate. *J. Oper. Oceanogr.* **1**, 9–18 (2008).

56. ASL Environmental Sciences. *Data Processing and Analysis of Ice Keel Depths, Fram Strait, 2006–2007. Report for Norsk Polarinstitutt, Tromsø, Norway* (ASL Environmental Sciences, 2009).

57. Hansen, E. et al. Thinning of Arctic sea ice observed in Fram Strait: 1990–2011. *J. Geophys. Res. Oceans* **118**, 5202–5221 (2013).

58. Huang, et al. Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *J. Clim.* **34**, 2923–2939 (2021).

59. Sumata, H. et al. Decorrelation scales for Arctic Ocean hydrography—Part I: Amerasian Basin. *Ocean Sci.* **14**, 161–185 (2018).

60. Sumata, H., Kauka, F., Karcher, M. & Rüdiger, G. Covariance of optimal parameters of an Arctic sea ice-ocean model. *Mon. Weather Rev.* **147**, 2579–2602 (2019).

61. Hersbach, H. et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).

62. Vinje, T. & Finnekåsa, Ø. The ice transport through Fram Strait. *Nor. Polarinst. Skr.* **186**, 1–41 (1986).

63. Künsch, H. R. The jackknife and the bootstrap for general stationary observations. *Ann. Stat.* **17**, 1217–1241 (1989).

64. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman & Hall, 1993).

65. Rodionov, S. & Overland, J. E. Application of a sequential regime shift detection methods to the Bering Sea ecosystem. *ICES J. Mar. Sci.* **62**, 328–332 (2005).

66. Mahoney, A. R., Hutchings, J. K., Eicken, H. & Haas, C. Changes in the thickness and circulation of multiyear ice in the Beaufort Gyre determined from pseudo-Lagrangian methods from 2003–2015. *J. Geophys. Res. Oceans* **124**, 5618–5633 (2019).

67. Kuninaka, H., Kobayashi, N. & Matsushita, M. Regularities hidden in complex systems—around the lognormal distribution [in Japanese]. *J. Phys. Soc. Jpn* **66**, 658–665 (2011).

68. Rothrock, D. A. The energetics of the plastic deformation of pack ice by ridging. *J. Geophys. Res.* **80**, 4514–4519 (1975).

69. Hopkins, M. A. On the ridging of intact lead ice. *J. Geophys. Res.* **99**, 16351–16360 (1994).

70. Itkin, P. et al. Contribution of deformation to sea ice mass balance: a case study from an N-ICE2015 storm. *Geophys. Res. Lett.* **45**, 789–796 (2018).

71. Valkonen, E., Cassano, J. & Cassano, E. Arctic cyclones and their interactions with the declining sea ice: a recent climatology. *J. Geophys. Res. Atmos.* **126**, e2020JD034366 (2021).

72. Leppäranta, M. A review of analytical models of sea-ice growth. *Atmos. Ocean* **31**, 123–138 (1993).

## Acknowledgements

This study has been made possible by the long-term observations from the Fram Strait Arctic Outflow Observatory maintained by the Norwegian Polar Institute. This work was supported by the Norwegian Research Council (grant no. 286971, project FreshArc) and partly received funding from the European Union's Horizon 2020 research and innovation programme under grant no. 101003826 via project CRiceS (Climate Relevant interactions and feedbacks: the key role of sea ice and Snow in the polar and global climate system). We thank J. Haapala and M. Lensu for discussions about the proportionate ice growth model.

## Author information

### Authors and Affiliations

1. Norwegian Polar Institute, Fram Centre, Tromsø, Norway

   Hiroshi Sumata, Laura de Steur, Dmitry V. Divine, Mats A. Granskog & Sebastian Gerland

### Contributions

H.S. and D.V.D. conceptualized the study. H.S., D.V.D. and M.A.G. devised the methodology. H.S. processed the data, devised the theoretical model, carried out the formal analysis and data visualization and wrote the original draft. H.S. and D.V.D. carried out the uncertainty assessments. H.S., L.d.S., D.V.D., S.G. and M.A.G. carried out the investigation. H.S., L.d.S., S.G., D.V.D. and M.A.G. reviewed and edited the manuscript. L.d.S. (FreshArc) and M.A.G. (CRiceS) acquired the funding. L.d.S. was the project lead and organized the study.

### Corresponding author

Correspondence to Hiroshi Sumata.

# Ethics declarations

## Competing interests

The authors declare no competing interests.

# Peer review

## Peer review information

*Nature* thanks Andrew Mahoney and Yusuke Kawaguchi for their contribution to the peer review of this work. Peer reviewer reports are available.

# Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# Extended data figures and tables

## Extended Data Fig. 1 Difference of sea surface temperature (SST) between two periods 1990–2006 and 2007–2019.

(a) Difference of mean September SST estimated from Daily Optimum Interpolated Sea Surface Temperature data set (DOISST ver. 2.1)[58], (b) Time series of mean September SST in sea ice formation areas A and B, calculated from DOISST. (c) Difference of upper ocean temperature (July to September mean, 0–20 m) between the two periods calculated from in-situ observational datasets[59,60]. The dashed lines in (b) denote detected regimes by sequential t-test described in Methods. Matplotlib basemap toolkit is used to plot the map.

## Extended Data Fig. 2 Mean residence time of sea ice in the Siberian sector and the central Arctic.

The residence time is calculated by the backward trajectories described in Methods. The central Arctic is defined being outside of the two polygons A and B. The ice formation areas A and B are shown in Fig. 3b in the main text. The solid lines denote regimes detected by the sequential t-test described in Methods.

## Extended Data Fig. 3 Time series of ice thickness distribution in Fram Strait including open water fraction.

The thickness distributions are derived including open water fraction (i.e., zero thickness bin) on monthly basis. Data processing procedures are described in Methods.

## Extended Data Fig. 4 Time series of ice thickness distribution in each site.

Time series of ice thickness distribution observed by each moored ULS (F11 to F14) in Fram Strait. The distributions are derived on monthly basis. Data processing procedures are described in Methods.

## Extended Data Fig. 5 Examples of winter and summer sea ice thickness distributions in Fram Strait.

Examples of sea ice thickness distributions and corresponding fitted lognormal functions in March (top three rows) and September (bottom three rows). The blue, orange and green lines show ice thickness distribution, fitted lognormal function, and cut-off threshold, respectively. The plots are shown for every three years if data are available. Data processing procedures are described in Methods.

## Extended Data Fig. 6 Time series of modal peak height and fitting parameters.

Time series of (a) modal thickness, (b) modal peak height, (c) variance, and (d, e) fitting parameters of lognormal functions. Data processing procedures are described in Methods. The gray solid lines in panels (a–c) show regimes detected by the sequential t-test.

## Extended Data Fig. 7 Change of sea level pressure and wind pattern after the regime shift.

Difference of sea level pressure (SLP) and wind field between two periods 1990–2006 and 2007–2019, in (a) summer (from June to November) and (b) winter (from December to May), and (c) time series of annual mean 10 m wind averaged in the three polygons shown in panels (a) and (b). The polygons show (a) Alaskan sector, (b) Siberian sector, and (c) the area representing the Transpolar Drift Stream. The mean speed in the rectangular box C is the component of 10 m wind vector parallel to the major axis of the box (positive wind speed orients to the Fram Strait). SLP and 10 m wind data are taken from ERA5[61]. Matplotlib basemap toolkit is used to plot the map.

## Extended Data Fig. 8 Fraction of thick sea ice with detected regimes.

Fraction of thick sea ice (> 5 m and > 4m) observed in Fram Strait. Data processing procedures are described in Methods. The solid and dashed lines denote regimes detected by the sequential *t*-test.

**Extended Data Table 1 Summary of sequential *t*-test analysis of the regime shift detection**
**Extended Data Table 2 Details of sequential *t*-test analysis of the regime shift detection**

# Supplementary information

## Peer Review File

# Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

# About this article

## Cite this article

Sumata, H., de Steur, L., Divine, D.V. *et al.* Regime shift in Arctic Ocean sea ice thickness. *Nature* **615**, 443–449 (2023). https://doi.org/10.1038/s41586-022-05686-x

- Received: 19 August 2022
- Accepted: 23 December 2022
- Published: 15 March 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-022-05686-x

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

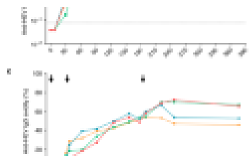
Provided by the Springer Nature SharedIt content-sharing initiative

## An abrupt decline of thick sea ice in the Arctic Ocean

Nature Research Briefing 15 Mar 2023

---

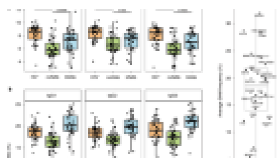
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-022-05686-x



## Quantitative evaluation of protective antibody response induced by hepatitis E vaccine in humans

07 August 2020

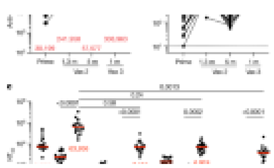
Gui-Ping Wen, Linling He, … Ning-Shao Xia



## Aberrant B cell repertoire selection associated with HIV neutralizing antibody breadth

20 January 2020

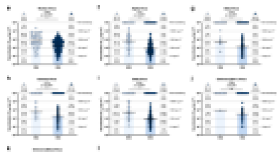
Krishna M. Roskin, Katherine J. L. Jackson, … Scott D. Boyd



## Increased memory B cell potency and breadth after a SARS-CoV-2 mRNA boost

21 April 2022

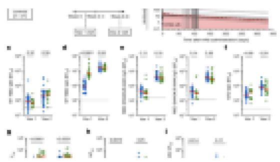
Frauke Muecksch, Zijun Wang, … Michel C. Nussenzweig



[B cell analyses after SARS-CoV-2 mRNA third vaccination reveals a hybrid immunity like antibody response]

04 January 2023

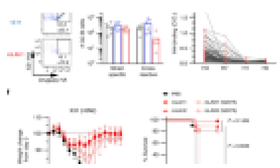
Emanuele Andreano, Ida Paciello, … Rino Rappuoli



[Antibody feedback regulates immune memory after SARS-CoV-2 mRNA vaccination]

06 December 2022

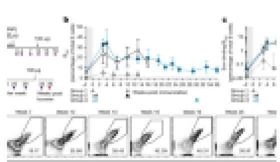
Dennis Schaefer-Babajew, Zijun Wang, … Michel C. Nussenzweig



[Exposure of an occluded hemagglutinin epitope drives selection of a class of cross-protective influenza antibodies]

28 August 2019

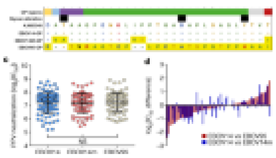
Yu Adachi, Keisuke Tonouchi, … Yoshimasa Takahashi



[**Long-primed germinal centres with enduring affinity maturation and clonal migration**]

21 September 2022

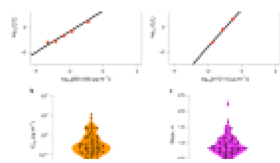
Jeong Hyun Lee, Henry J. Sutton, … Shane Crotty



[**Ebola virus antibody decay–stimulation in a high proportion of survivors**]

27 January 2021

Charlene Adaken, Janet T. Scott, … Georgios Pollakis



[**Modeling how antibody responses may determine the efficacy of COVID-19 vaccines**]

28 February 2022

Pranesh Padmanabhan, Rajat Desikan & Narendra M. Dixit

- Article
- Published: 16 January 2023

# Molecular fate-mapping of serum antibody responses to repeat

# immunization


- Ariën Schiepers,
- Marije F. L. van 't Wout,
- Allison J. Greaney,
- Trinity Zang,
- Hiromi Muramatsu,
- Paulo J. C. Lin,
- Ying K. Tam,
- Luka Mesin,
- Tyler N. Starr,
- Paul D. Bieniasz,
- Norbert Pardi,
- Jesse D. Bloom &
- …
- Gabriel D. Victora

Show authors



*Nature* **volume 615**, pages 482–489 (2023)



- 27k Accesses
- 2 Citations
- 207 Altmetric
- Metrics details


## Abstract


The protective efficacy of serum antibodies results from the interplay of antigen-specific B cell clones of different affinities and specificities. These cellular dynamics underlie serum-level phenomena such as original antigenic sin (OAS)—a proposed propensity of the immune system to rely repeatedly on the first cohort of B cells engaged by an antigenic stimulus

when encountering related antigens, in detriment to the induction of de novo responses[1,2,3,4,5]. OAS-type suppression of new, variant-specific antibodies may pose a barrier to vaccination against rapidly evolving viruses such as influenza and SARS-CoV-2[6,7]. Precise measurement of OAS-type suppression is challenging because cellular and temporal origins cannot readily be ascribed to antibodies in circulation; its effect on subsequent antibody responses therefore remains unclear[5,8]. Here we introduce a molecular fate-mapping approach with which serum antibodies derived from specific cohorts of B cells can be differentially detected. We show that serum responses to sequential homologous boosting derive overwhelmingly from primary cohort B cells, while later induction of new antibody responses from naive B cells is strongly suppressed. Such 'primary addiction' decreases sharply as a function of antigenic distance, allowing reimmunization with divergent viral glycoproteins to produce de novo antibody responses targeting epitopes that are absent from the priming variant. Our findings have implications for the understanding of OAS and for the design and testing of vaccines against evolving pathogens.

# Main

The ability of serum antibodies to protect against infection is an emergent property of the complex mixture of immunoglobulins secreted over time by B cell clones of various specificities and of a range of affinities. The plasma cells that produce these antibodies arise through multiple parallel pathways, ranging from direct differentiation from naive B cell precursors after primary infection or immunization to elaborate routes involving one or more rounds of affinity maturation in germinal centres (GCs) and intercalating memory B cell phases. The complexity of these cellular pathways compounds markedly with repeated antigenic exposure[9,10,11,12], and their ultimate contribution to the serum antibody pool has been difficult to deconvolute. On the one hand, molecular analyses of immunoglobulin genes obtained from memory or GC B cells do not directly assess the composition of antibodies in the serum[13,14,15,16]; on the other hand, direct studies of the clonal composition of serum antibody cannot readily assign a cellular or temporal origin to antibodies of different specificities[17,18]. The clonal

dynamics of immune phenomena that take place at the serum level therefore remain poorly understood.

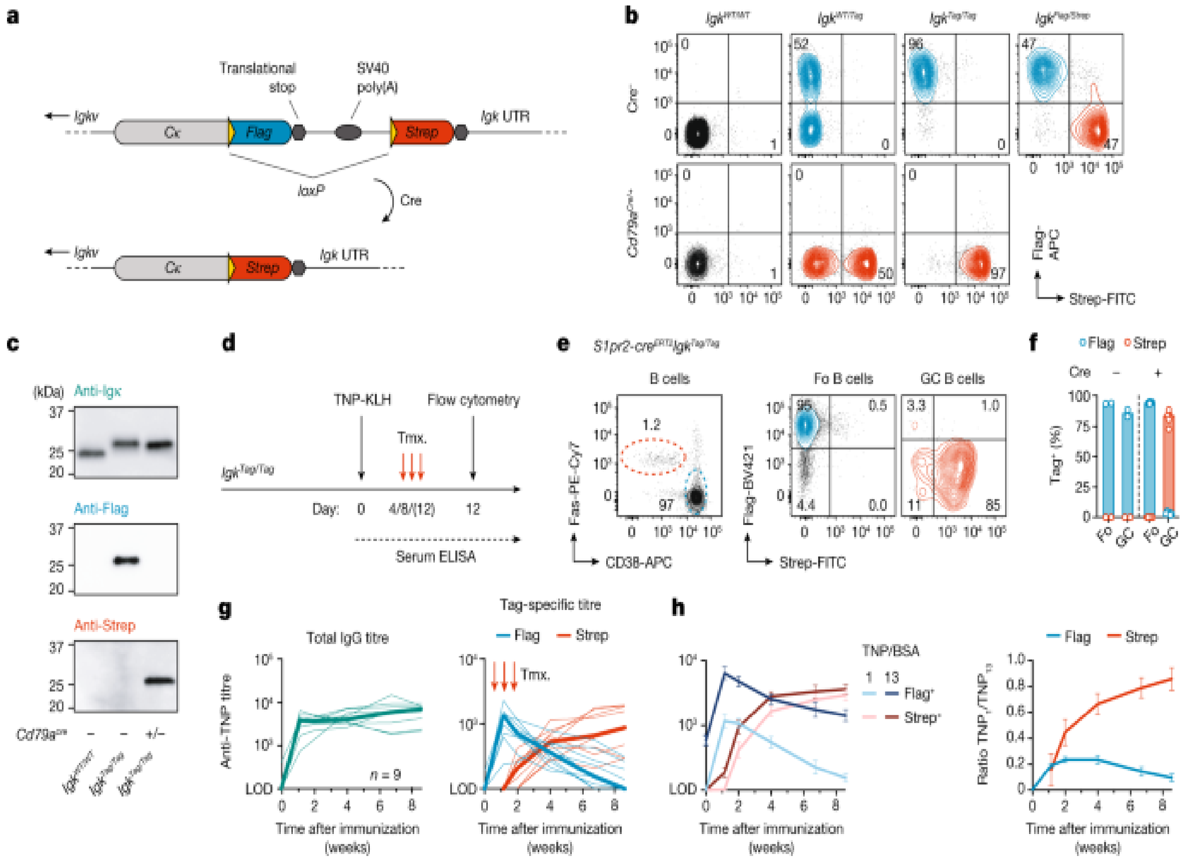
A serum-level phenomenon that has been particularly difficult to unravel is OAS, described in the 1950s as a tendency of individuals exposed to a given strain of influenza to respond with antibodies that react more strongly to the first strain of influenza they had met in early childhood than to the exposure strain itself[1,19]. OAS was originally attributed to a propensity of the immune system to repeatedly reuse the first cohort of B cells that respond to an antigen, the reactivity of which will necessarily be biased towards the strain that originally elicited it. However, in contrast to related concepts such as antigenic seniority[4,20], OAS (as defined herein) requires active suppression of the de novo recruitment of new B cell clones from the naive repertoire after boosting[2,3,4], therefore restricting the ability of the immune system to mount specific antibody responses to escape epitopes. The extent to which this active suppression exists and influences subsequent responses has been debated for decades[5,8]. More recently, B cell fate-mapping experiments in mice have shown that, in apparent contrast to the predictions of OAS, GCs that form in response to boosting consist almost exclusively of naive rather than memory-derived B cells[21,22,23]. This later addition implies that either the effects of OAS in mice are negligible, or that OAS is a phenomenon that is observed exclusively at the serum level.

# Molecular fate-mapping of serum antibody

Resolving this issue, as well as generally understanding the effect of OAS on the response to repeated antigen exposure, would require the ability to transpose such cellular fate-mapping experiments to the serum antibody itself. To achieve this, we adapted the classic fate-mapping strategy to enable the detection of the cellular and temporal origin of antibodies in the serum—an approach that we call molecular fate-mapping. We engineered mice in which the C terminus of the immunoglobulin kappa (Igκ) light chain gene (*Igk*) is extended to encode a LoxP-flanked Flag tag, followed by a downstream Strep tag (Fig. 1a and Extended Data Fig. 1). B cells bearing this ‘K-tag’ allele produce immunoglobulins that are Flag-tagged unless they are exposed to Cre recombinase, after which they permanently switch the Flag tag for a Strep tag. Cre-mediated recombination therefore fate-maps the

antibodies that these B cells and their plasma cell descendants express on their surfaces and/or secrete into serum. This enables differential detection of pre- and post-fate-mapping Igκ$^{+}$ antibodies using secondary reagents specific for each tag.

**Fig. 1: The K-tag system for molecular fate-mapping of serum antibodies.**




**a**, Schematic of the *Igk*$^{Tag}$ (K-tag) allele before and after Cre-mediated recombination. UTR, untranslated region. **b**, Flow cytometry analysis of blood B cells showing the expression of Flag- and Strep-tagged B cell receptors on mice of the indicated genotype. **c**, Western blot analysis of serum obtained from mice of the indicated genotype, stained for Igκ light chain or Flag/Strep tags. Representative of two experiments. **d**, Schematic of the immunization strategy used to fate-map GC B cells and their antibody output. **e**, Flow cytometry analysis of the popliteal lymph node 12 days after footpad immunization with TNP-KLH in alum adjuvant. B cells ($B220^{+}CD4^{-}CD8^{-}CD138^{-}$) were stained for GC ($Fas^{+}CD38^{-}$) and follicular (Fo) B cell ($Fas^{-}CD38^{+}$) markers. **f**, Quantification of the data in **e**. Each dot

represents an individual lymph node, and the bars represent the median values. **g**, Anti-TNP total IgG and tag-specific end-point titres as determined by $TNP_4$-BSA ELISA in mice immunized i.p. with TNP-KLH in alhydrogel adjuvant. The thin lines represent individual mice and the thick lines link the medians of log-transformed titre values at each time point. Results are from nine mice from two independent experiments. Tmx., tamoxifen. **h**, The relative affinity of anti-TNP $Flag^+$ and $Strep^+$ antibodies of the same samples shown in **g** as estimated by ELISA using $TNP_1$-BSA or $TNP_{13}$-BSA as capture reagents. Data are mean ± s.e.m. of the log-transformed titre values. The ratio between anti-$TNP_1$-BSA and anti-$TNP_{13}$-BSA titres was calculated per sample, shown on the right.

To verify the functionality of the K-tag allele, we first determined that B cells in *Igk*$^{Tag}$ mice expressed tagged B cell receptors on their surface. Following the rules of allelic exclusion[24], around 50% and 95% of B cells from mice expressing heterozygous *Igk* (wild type (WT)/tag) or homozygous *Igk*$^{Tag/Tag}$ mice were $Flag^+$ (as expected, about 5% of B cells in homozygous mice carried an Igλ light chain[24]; Extended Data Fig. 2a). K-tag mice in which all B cells constitutively expressed Cre recombinase (*Igk*$^{Tag/Tag}$*Cd79a*$^{Cre/+}$) replaced Flag with a Strep tag in almost all $Ig\kappa^+$ B cells (Fig. 1b). Importantly, K-tag mice appropriately secreted Flag- or Strep-tagged antibodies into the serum in the absence or presence of Cre recombinase, respectively (Fig. 1c), without affecting steady-state serum antibody levels (Extended Data Fig. 2b). The generation and maturation of K-tag B cells was unimpaired, as indicated by the equal proportion of tagged and untagged circulating B cells in *Igk*$^{WT/Tag}$ mice (Fig. 1b). The same was true of circulating B cells and bone marrow plasma cells expressing Flag and Strep tags in *Igk*$^{Flag/Strep}$mice, in which one of the two K-tag alleles was prerecombined by Cre expression in the germline (Fig. 1b and Extended Data Fig. 2c).

To ensure that $Flag^+$ and $Strep^+$ B cells were equally competitive throughout the course of B cell activation, affinity maturation and plasma cell differentiation, we primed and boosted *Igk*$^{Flag/Strep}$ mice with the model antigen 2,4,6-trinitrophenyl-keyhole limpet haemocyanin (TNP-KLH) in alum adjuvant and followed the serum titres of antibodies bearing each tag

over time. To enable a direct comparison of titres of anti-TNP antibodies bearing each tag, we diluted secondary (anti-Flag or anti-Strep) antibodies to achieve similar detection of standard curves generated using recombinant Flag- or Strep-tagged monoclonal antibodies (Extended Data Fig. 2d). End-point enzyme-linked immunosorbent assay (ELISA) titres normalized using these curves showed a similar range of anti-TNP reactivity between the $Flag^{+}$ and $Strep^{+}$ fractions (Extended Data Fig. 2e), indicative of equal competitiveness of the differently tagged B cells.

Following the serum antibody produced by B cells engaged at different stages of the immune response requires temporally restricted fate-mapping of activated B cell clones. To enable this, we crossed $Igk^{Tag}$ mice to the GC-specific, tamoxifen-inducible $S1pr2\text{-}cre^{ERT2}$ BAC-transgenic allele[25] (to generate $S1pr2\text{-}Igk^{Tag}$ mice). Tamoxifen treatment of mice on days 4 and 8 after TNP-KLH immunization led to efficient recombination (96.1 ± 0.50% (mean ± s.e.m.) (($Strep^{+}/Tag^{+}$) × 100)) of the K-tag allele in GC B cells but not in non-GC B cells in the same lymph node at 12 days after immunization (d.p.i.) (Fig. 1d–f). Again, tagged B cells were found at similar proportions to untagged B cells in heterozygous $S1pr2\text{-}Igk^{WT/Tag}$ mice (on average 41 ± 9.0% (mean ± s.d.) $Tag^{+}$), indicating that expression of the tag does not impair B cell competitiveness in the GC (Extended Data Fig. 2f,g). GC B cells in $Cre^{-}$ animals (Fig. 1f) or $S1pr2\text{-}Igk^{WT/Tag}$ mice not treated with tamoxifen remained $Flag^{+}$, with only minimal spontaneous recombination (1.3 ± 1.0% (mean ± s.d.) $Strep^{+}$ at day 12 d.p.i.) in the latter (Extended Data Fig. 2g).

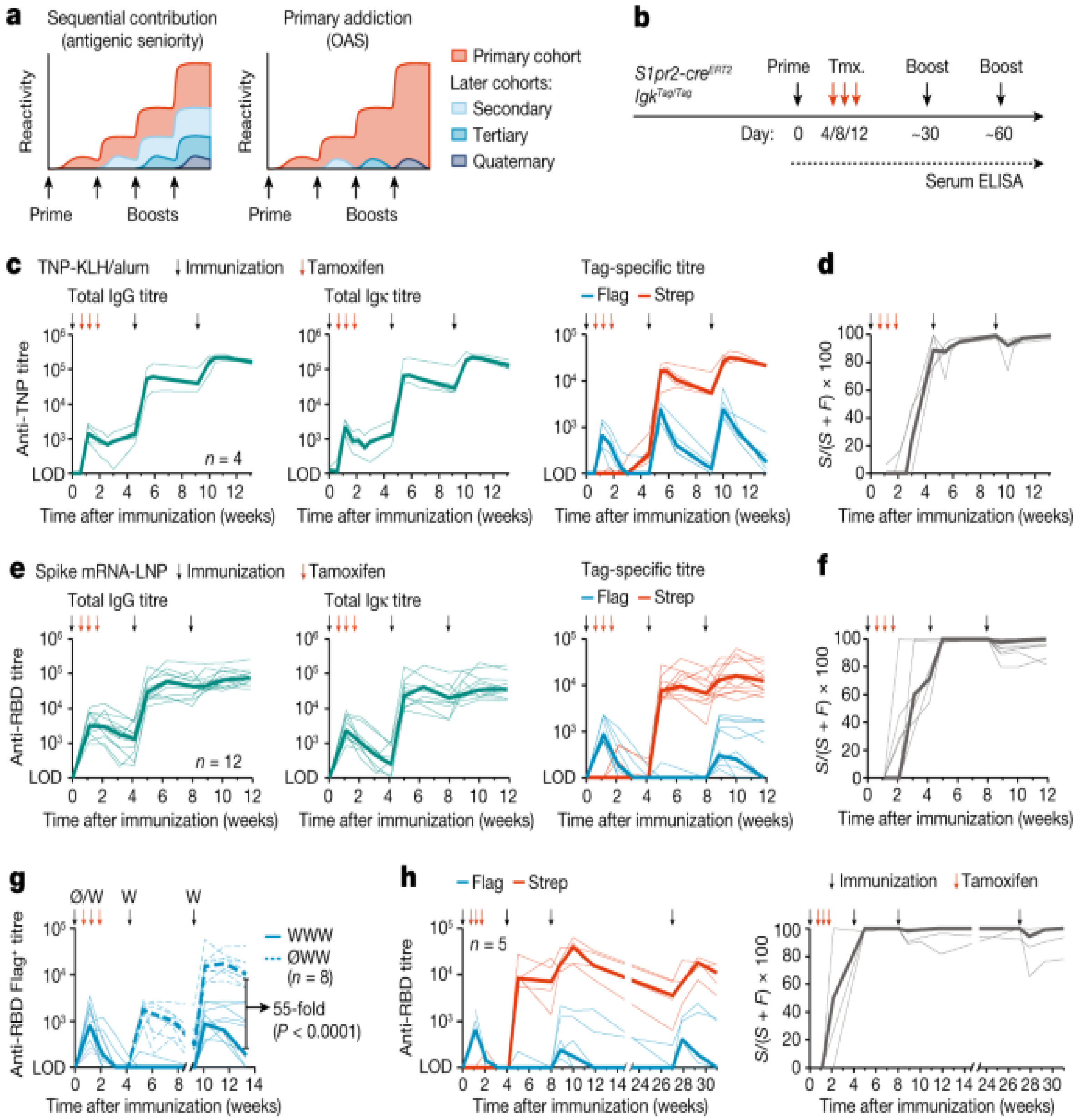
Total anti-TNP IgG antibodies in $S1pr2\text{-}Igk^{Tag}$ mice immunized intraperitoneally (i.p.) with TNP-KLH in alhydrogel and treated with tamoxifen on days 4, 8 and 12 were first detected in the serum at 8 d.p.i. and increased progressively until 60 d.p.i. (Fig. 1g). Deconvolution of GC-derived ($Strep^{+}$) and non-GC derived ($Flag^{+}$) antibodies showed that an initial wave of extrafollicular $Flag^{+}$ antibodies that peaked at 8 d.p.i. was progressively replaced by GC-derived $Strep^{+}$ antibodies that were first detected in the serum at 14 d.p.i. (Fig. 1g). $Flag^{+}$ anti-TNP antibodies regressed to near baseline levels between 47–60 d.p.i., as expected from their extrafollicular origin. An affinity-dependent anti-TNP ELISA showed

detectable affinity maturation only in the GC-derived Strep$^{+}$ antibody fraction (Fig. 1h), confirming the efficient fate-mapping of GC-derived antibody. The background signal in control animals that were not given tamoxifen remained below the limit of detection (LOD) throughout the primary response (Extended Data Fig. 2h). Thus, the *S1pr2-Igk$^{Tag}$* mouse model enables us to discern antibodies derived from the first wave of B cells that entered a GC reaction in response to immunization. Our data also show that extrafollicular responses are of relatively short duration in these settings and that almost all antibodies detectable after the first few weeks of immunization, and especially antibodies with high affinity, were derived from plasma cells of GC origin.

## Primary addiction in homologous boosting

With this system in hand, we sought to measure the extent to which OAS-type suppression affects the development of de novo antibody responses to homologous boosting (we refer to this suppression generically as primary addiction, to encompass the homologous regimen). To this end, we took advantage of the ability to trigger Cre-mediated recombination of the K-tag allele in a time-resolved manner by administration of tamoxifen to mark serum antibodies produced by B cells that formed GCs in response to primary immunization (the primary cohort). In addition to labelling primary-cohort B cells, this approach also ‘reverse fate-maps’ with a Flag tag any antibodies that arose from clones engaged de novo by subsequent booster doses. This property enables us to distinguish between two models of recall antibody response: (1) a simple sequential contribution model, in which de novo responses, although smaller than memory-derived ones, are nevertheless allowed to progress with similar kinetics to a new primary response, adding up as more doses of antigen are provided (related to antigenic seniority); and (2) a primary addiction model (related to OAS), in which the primary response actively suppresses the emergence of subsequent de novo antibody responses even after several boosts (Fig. 2a).

**Fig. 2: Primary addiction after homologous boosting.**




**a**, Schematic of the sequential contribution and primary addiction models of antigenic imprinting. **b**, The general immunization strategy used to measure primary addiction. *S1pr2-Igk*$^{Tag/Tag}$ mice were immunized on the days indicated by black arrows with TNP-KLH in alum (**c**,**d**) or WH1 mRNA–LNP (**e**–**g**) and treated with tamoxifen at 4, 8 and 12 d.p.i. as indicated by the red arrows. **c**, Anti-TNP serum IgG (left), Igκ (middle) and tag-specific titres (right) as measured using $TNP_4$-BSA by ELISA. The results are from four mice from two independent experiments. The thin lines represent individual mice and the thick lines link the medians of log-transformed titre

values at each timepoint. **d**, The percentage of the TNP titre derived from the primary cohort of B cells (the primary addiction index) was calculated by dividing the Strep$^+$ titre of each individual sample by its total titre (Strep$^+$ + Flag$^+$), multiplied by 100 ($S/(S+F)\times 100$). **e**, Anti-WH1 RBD IgG (left), Igκ (middle) and tag-specific titres (right) as measured by ELISA. Results are from 12 mice from 3 independent experiments. **f**, The primary addiction index was calculated as described in **d**. **g**, Comparison of de novo Flag$^+$ antibody responses in the presence or absence of primary immunization. WWW, three doses of WH1 S mRNA; ØWW, first dose and fate-mapping omitted. WWW data are for the same samples as in **e**, remeasured in the same assay as ØWW. **h**, The quaternary anti-WH1 RBD response (left) and primary addiction index (right) in mice that received a fourth dose of mRNA–LNP at 133 days after the previous dose, for one of the cohorts shown in **e**. Two out of the five mice were not sampled at day 0.

We primed *S1pr2-Igk$^{Tag}$* mice i.p. with alum-adjuvanted TNP-KLH and administered tamoxifen at 4, 8 and 12 d.p.i. to fate-map the primary cohort GC B cells and their memory and plasma cell progeny (Fig. 2b). In this setting, all primary-cohort-derived antibodies are Strep$^+$, whereas any antibody produced by B cell clones expanded de novo by secondary or higher-order boosting (including by their memory or plasma cell progeny) will be reverse fate-mapped as Flag$^+$. Importantly, as we do not rely on differences between antigen variants to distinguish primary from secondary or later antibodies, this approach enables us to measure primary addiction at zero-antigenic distance—that is, when priming and boosting with the exact same antigen. As suppression of de novo antibody responses is likely to decrease as antigenic distance increases[26,27], this approach enables us to estimate the strength of primary addiction when it is at its strongest.

Homologous boosting 1 and 2 months after the primary immunization resulted in the expected increases in recall TNP titres (Fig. 2c) and the formation of recall GCs that were dominated by naive-derived B cells[21] (Extended Data Fig. 3a). Deconvolution of these responses using tag-specific ELISA revealed that both secondary and tertiary titres were strongly dominated by fate-mapped (Strep$^+$) antibodies derived from primary cohort B cells. Whereas Flag$^+$ TNP-specific antibodies also appeared after each boost, their titres peaked at much lower levels and decayed markedly with

time (Fig. 2c). Importantly, counter to the expectations of the sequential contribution model (Fig. 2a), $Flag^{+}$ titres did not increase progressively between the second and third antigen doses, when any $Flag^{+}$ memory B cells would have been reactivated. To synthesize both measures, we created a 'primary addiction index', computed by dividing $Strep^{+}$ by total $Strep^{+}$ + $Flag^{+}$ titres ($S/(S+F) \times 100$). This showed that almost all detectable recall antibodies (mean 95% and 97% of serum reactivity at 14 days after the first and second boosts, respectively) were derived from the B cell cohort engaged in the primary GC response (Fig. 2d). Depletion of IgM from serum samples after boosting resulted in a sharp reduction in $Flag^{+}$ but not $Strep^{+}$ recall TNP titres, supporting the notion that naive-derived B cells engaged by recall generate primarily an extrafollicular-like (IgM-dominated) B cell response (Extended Data Fig. 3b,c).

To extend these findings to a clinically relevant setting, we immunized and boosted mice as in Fig. 2b but using a lipid nanoparticle (LNP)-formulated nucleoside-modified mRNA vaccine encoding the prefusion-stabilized (2P) form of the SARS-CoV-2 Wuhan-Hu-1 (WH1) spike (S) protein, similar to available SARS-CoV-2 mRNA vaccines[28]. Secondary and tertiary anti-S protein receptor-binding domain (RBD) antibodies were again almost entirely derived from primary cohort ($Strep^{+}$) B cells (Fig. 2e,f). As with GC B cells (Extended Data Fig. 2g), low-level spontaneous recombination to $Strep^{+}$ antibodies was detected in recall responses in control mice not given tamoxifen. This resulted in a slight underestimation of $Flag^{+}$ antibody titres (median 2.1% and 2.8% 2 weeks after second and third immunizations, respectively; Extended Data Fig. 3d). Although primary addiction was more pronounced for the SARS-CoV-2 RBD than for TNP-KLH after the first boost (no new ($Flag^{+}$) antibody was detected at this time point), 5 out of 12 mice developed low but stable titres of $Flag^{+}$ anti-RBD antibodies after the third dose. This bimodality was independent of the experimental cohort and of whether boosting was done ipsilaterally or contralaterally to the site of the primary dose (Extended Data Fig. 3e) and is therefore likely ascribable to stochastic variability inherent to highly oligoclonal recall responses[21]. To quantify the extent to which de novo antibody responses to the boost were suppressed by previous priming (that is, the magnitude of the primary addiction suppressive effect), we compared $Flag^{+}$ antibodies in mice that

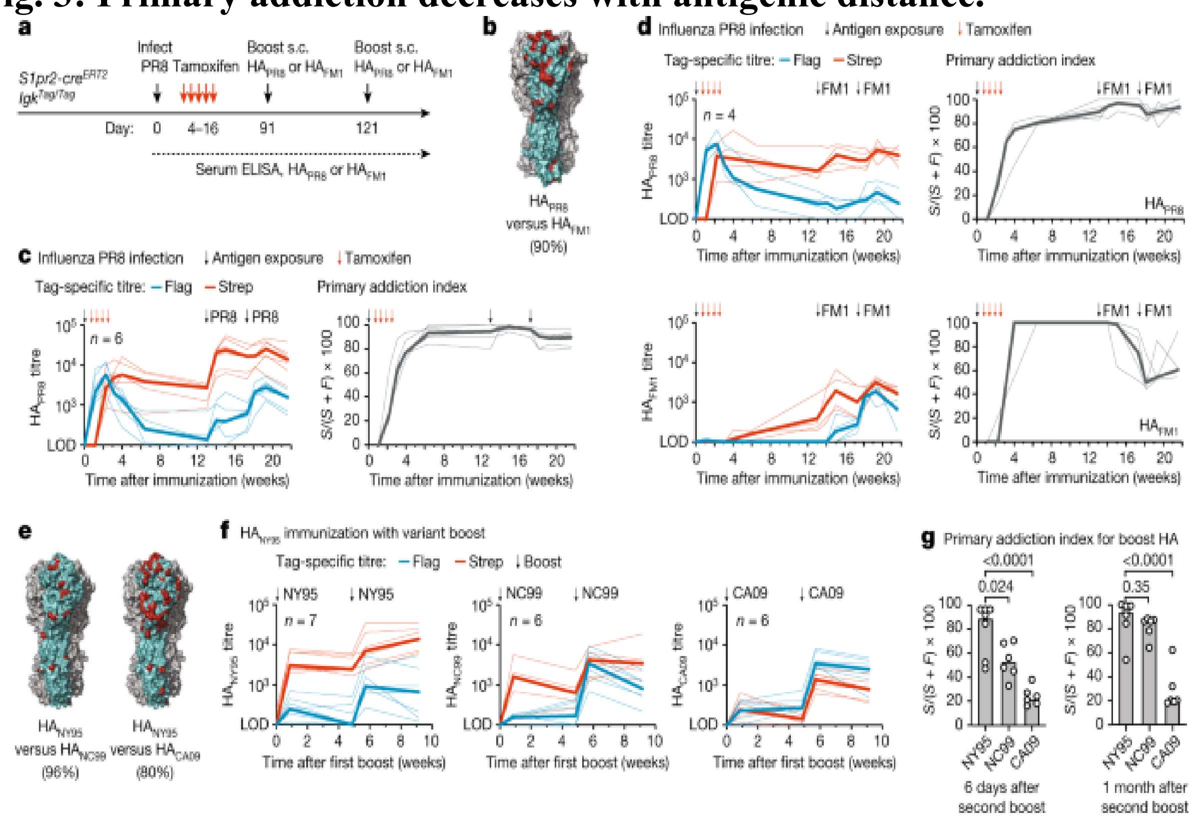
were given three doses of WH1 mRNA–LNPs (WWW) to an additional group in which the priming and fate-mapping steps were omitted (ØWW). $Flag^+$ responses were 55-fold lower in WWW mice compared with in ØWW mice at 4 weeks after the final dose (Fig. 2g), indicating strong suppression of new B cell responses in primed animals compared with what they would have been in the absence of priming. Finally, primary addiction was long lasting, as even a fourth immunization of a subset of mice (>133 days after the previous boost) was dominated by Strep-tagged antibodies (Fig. 2h and Extended Data Fig. 3f), again failing to demonstrate the progressive increase in $Flag^+$ antibody titres predicted by a simple sequential contribution model (Fig. 2a). We conclude that primary addiction can be very strong when measured at zero antigenic distance—evidence of OAS-type suppression of de novo B cell responses by pre-existing immunity.

# Antigenic drift limits primary addiction

To measure how primary addiction responds to increases in antigenic distance between priming and boosting antigens, we used historical series of drifted influenza virus haemagglutinin (HA) variants as models. We first used an influenza infection/immunization model (Fig. 3a) based on two of the strains for which OAS was originally described—A/Puerto Rico/8/1934 (PR8) and A/Fort Monmouth/1/1947 (FM1)[1,19]—of which the HAs ($HA_{PR8}$ and $HA_{FM1}$) share 90% identity at the amino acid level (Fig. 3b). As with hapten and mRNA immunization, the primary response to $HA_{PR8}$ was characterized by high extrafollicular ($Flag^+$) titres that peaked between 8 and 16 days after infection and were subsequently replaced by GC-derived ($Strep^+$) titres (Fig. 3c). Homologous boosting with recombinant $HA_{PR8}$ protein subcutaneously at 3 and 4 months after infection resulted in a 1 log increase in $Strep^+$ $HA_{PR8}$ binding titres after the first boost and a less pronounced increase after the second boost. As with protein immunization, the contribution of non-primary ($Flag^+$) antibodies to total titres was small—even though it increased progressively between the first and second boosts, its peak median value was around 10% of HA reactivity (Fig. 3c). Heterologous boosting with $HA_{FM1}$ led to only slight back-boosting of primary $Strep^+$ $HA_{PR8}$ titres and had almost no effect on $Flag^+$ $HA_{PR8}$

reactivity (Fig. 3d), indicative of substantial antigenic distance between these variants. Accordingly, cross-reactive primary titres towards $HA_{FM1}$ were completely absent from the primary extrafollicular response to PR8 infection and began to emerge only at approximately 4 weeks in the $Strep^{+}$ antibody fraction (Fig. 3d), probably as a side effect of affinity maturation towards $HA_{PR8}$. Heterologous boosting not only increased these cross-reactive ($Strep^{+}$) titres by close to 1 log but, importantly, also induced substantial responses from de novo clones elicited by the boost, in that around half of all serum reactivity to $HA_{FM1}$ was derived from the $Flag^{+}$ fraction after the second boost (Fig. 3d). Comparing these levels to those achieved by two doses of $HA_{FM1}$ in the absence of previous infection showed that primary addiction suppressed new responses by 3.8-fold (Extended Data Fig. 4a), much less than the 55-fold suppression achieved in homologous mRNA-vaccination (Fig. 2g). Thus, heterologous boosting partly circumvents primary addiction, enabling improved expansion and serum contribution of variant-specific B cell clones that are not involved in the primary response.

**Fig. 3: Primary addiction decreases with antigenic distance.**



**a**, Schematic of the influenza infection and HA boosting strategy. *S1pr2-Igk*$^{Tag/Tag}$ mice were infected intranasally with PR8 influenza and boosted subcutaneously (s.c.) with $HA_{PR8}$ or $HA_{FM1}$ in alhydrogel at the indicated time points. **b**, Rendering of the $HA_{PR8}$ trimer structure (Protein Data Bank (PDB): 1RU7) with one monomer highlighted in teal and the amino acids that diverge between $HA_{PR8}$ and $HA_{FM1}$ in red. The percentage amino acid identity is shown in parentheses. **c**, Anti-$HA_{PR8}$ $Flag^{+}$ and $Strep^{+}$ titres in *S1pr2-Igk*$^{Tag/Tag}$ mice that were homologously boosted with $HA_{PR8}$ (left) and quantification of the primary addiction index score (right). **d**, Anti-$HA_{PR8}$ (top) and anti-$HA_{FM1}$ (bottom) ELISA reactivity in mice that were heterologously boosted with $HA_{FM1}$. Tag-specific titres (left) and quantification of the primary addiction index (right) are shown. **e**, The divergence between $HA_{NC95}$ and $HA_{NC99}$ or $HA_{CA09}$, coloured as in **b**. Modelled on the structure of $HA_{CA09}$ (PDB: 3LZG). **f**, Anti-HA tag-specific titres in mice primed with $HA_{NY95}$ protein, shown after the first boost with $HA_{NY95}$ (homologous) or with variants $HA_{NC99}$ or $HA_{CA09}$ (heterologous), as outlined in Extended Data Fig. 4b. Antibody reactivity to the boosting antigen is shown. The full time-course and reactivities against all three HAs as measured by ELISA for the top serum dilution are shown in Extended Data Fig. 4c. **g**, Primary addiction index for the boosting HA, 6 days (left) and 1 month (right) after the second boost (third dose). *P* values were calculated using two-tailed Student's *t*-tests, comparing the primary addiction index of the homologous to each heterologous boost. The thin lines represent individual mice and the thick lines link medians of log-transformed titre values at each time point. All of the results are from two independent experiments, with the number of mice in each group indicated in the graphs.

To verify this notion over a wider range of antigenic distances, we immunized mice i.p. with recombinant H1 from strain A/New York/614/1995 ($HA_{NY95}$) in alhydrogel adjuvant, then boosted these mice twice, either homologously with $HA_{NY95}$ or heterologously with H1s from strains A/New Caledonia/20/1999 ($HA_{NC99}$; a slightly drifted strain with 96% amino acid identity to $HA_{NY95}$) or pandemic A/California/07/2009

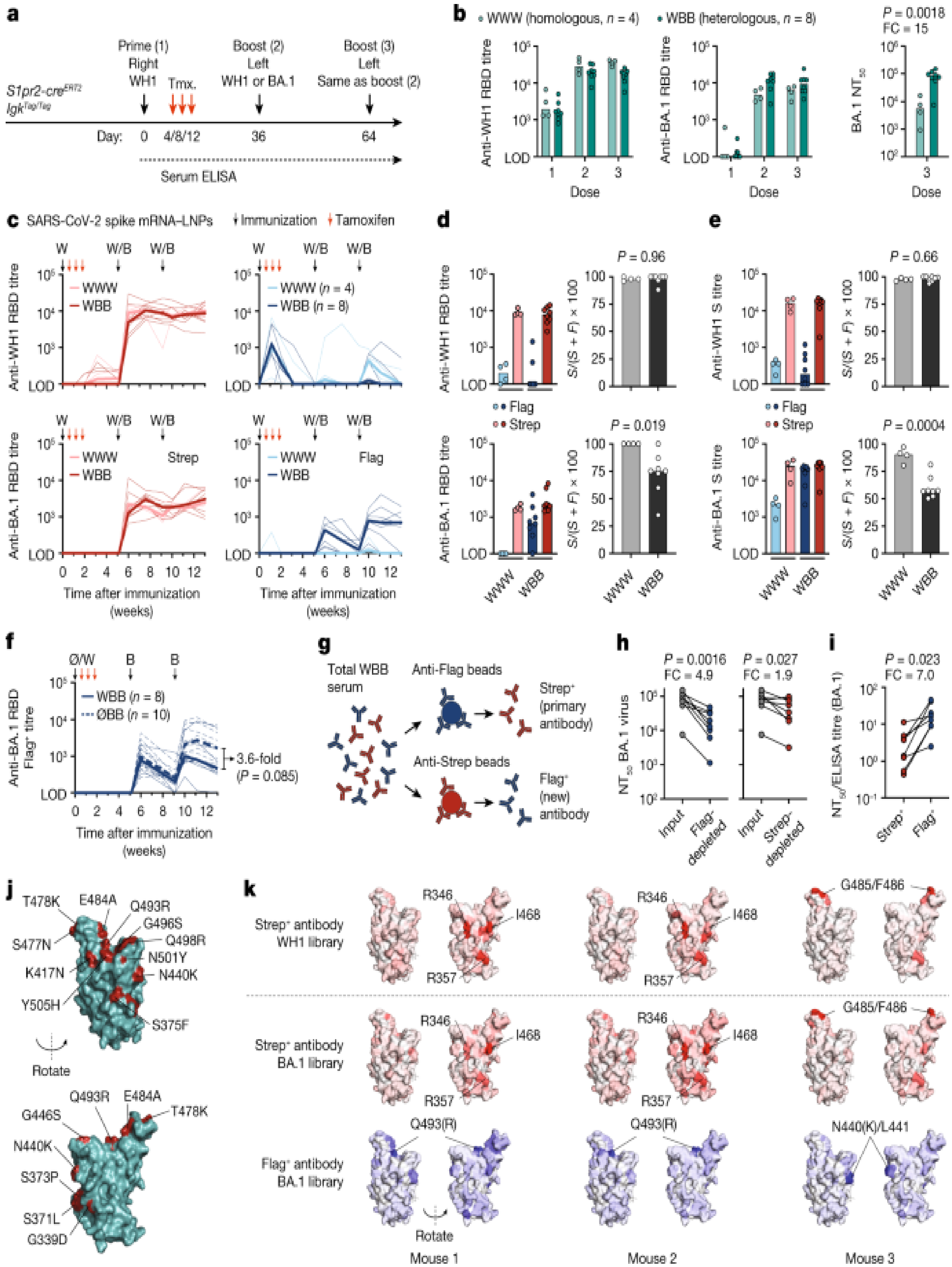
($HA_{CA09}$; an ‘antigenic shift’ strain, with 80% amino acid identity; Fig. 3e and Extended Data Fig. 4b). Generally, primary addiction was weaker and more variable in this setting, even with homologous boosting, possibly due to the overall weak primary response elicited by recombinant $HA_{PR8}$ protein (Extended Data Fig. 4c). Nevertheless, we observed a progressive decrease in primary addiction as the antigenic distance between the primary and boost antigens increased, so that up to 80% of total serum responses to $HA_{CA09}$ were Flag-tagged (corresponding to 20% primary addiction) after boosting with this variant (Fig. 3f,g). Pooling data for the infection and immunization experiments according to the similarity between priming and boosting HAs showed a highly significant linear decrease in primary addiction as antigenic distance increased (Extended Data Fig. 4d). We conclude that increased antigenic distance between priming and boosting antigens counteracts primary addiction, therefore enabling the generation of new, variant-specific antibody responses.

# Heterologous SARS-CoV-2 spike boosting

A setting in which subversion of primary addiction by antigenic distance is clinically important is the response to Omicron strains of SARS-CoV-2 in individuals who were previously exposed to antigens from the WH1 strain. We used the K-tag system to estimate the degree to which boosting with mRNA–LNP encoding the S protein from the Omicron BA.1 strain was capable of overcoming primary addiction generated by priming with WH1-S-encoding mRNA–LNP (the WH1 and BA.1 strains have 98% and 92% amino acid identity in the full S protein and RBD domains, respectively). We primed *S1pr2-Igk*$^{Tag}$ mice with WH1 mRNA–LNP in the right leg, then boosted these mice 1 and 2 months later with either BA.1 or WH1 mRNA–LNP distally in the left leg (Fig. 4a). Boosting induced similar total IgG responses to WH1 and BA.1 RBDs in both groups (Fig. 4b). By contrast, whereas sera from both groups neutralized a WH1 pseudovirus[29] equally, BA.1-boosted serum was on average 15-fold more potent against BA.1 pseudovirus, indicating a strong qualitative difference between the heterologous and homologous boosting regimens. Deconvolution of these effects by tag-specific ELISA revealed responses to the WH1 RBD that were indistinguishable between homologously and heterologously boosted

animals, in that primary ($Strep^+$) antibodies were strongly dominant in both settings, with no substantial $Flag^+$ response after the initial extrafollicular response (Fig. 4c,d). Whereas little to no $Strep^+$ antibodies to the BA.1 RBD were observed after primary immunization, recall $Strep^+$ reactivity to the BA.1 RBD was equally strong regardless of which variant was used for boosting (Fig. 4c,d). This observation agrees with previous reports documenting the evolution of cross-reactivity to other strains as a consequence of affinity maturation towards WH1 vaccination in humans[30]. Importantly, however, heterologous boosting resulted in a pronounced increase in BA.1 RBD titres generated by newly recruited ($Flag^+$) clones that were not cross-reactive with the WH1 strain, a reactivity otherwise absent from mice boosted homologously (Fig. 4c,d). At their peak (2 weeks after the second boost), $Strep^+$ antibodies accounted for an average of 73% (±19% s.d.) of total anti-BA.1 reactivity across heterologously boosted mice (Fig. 4d). The induction of new ($Flag^+$) antibodies after double BA.1 boost was even more pronounced when assaying for reactivity against the full-length WH1 and BA.1 S proteins (Fig. 4e). Two doses of BA.1 in the absence of previous WH1 immunization (ØBB) generated responses that were only 3.6-fold higher than the WBB (WH1-BA.1-BA.1) group (a difference that did not reach statistical significance; Fig. 4f), again showing that the effect of primary addiction in this setting is greatly reduced compared with that observed for homologous boosting (Fig. 2g). We conclude that BA.1 is sufficiently divergent from WH1 to induce substantial de novo antibody responses, even if it is not able to entirely overcome primary addiction. Moreover, the key difference between boosting homologously and heterologously is that only the latter can elicit a robust de novo response to the drifted strain.

**Fig. 4: Subversion of primary addiction by heterologous SARS-CoV-2 boosting.**



**a**, Schematic of the immunization strategies. **b**, Anti-WH1 (left) and BA.1 RBD IgG titres (middle) and $NT_{50}$ of BA.1 pseudovirus (right) 2 weeks after

the indicated dose in homologously (WWW) or heterologously (WBB) immunized *S1pr2-Igk*$^{Tag/Tag}$ mice. *P* values were calculated using two-tailed *t*-tests. FC, fold change. **c**, Evolution of anti-WH1 and BA.1 RBD tag-specific titres. The thin lines represent individual mice and the thick lines link the medians of log-transformed titre values. **d**, Comparison of the Flag and Strep anti-RBD titres shown in **c** at 2 weeks after the third immunization (left) along with primary addiction index (right). *P* values were calculated using two-tailed *t*-tests. **e**, Anti-full-S tag-specific titres and the primary addiction index for the same samples as in **d**, 2 weeks after the third dose. The results in **b**–**d** are from two independent experiments; the number of mice in each group is indicated. **f**, Comparison of de novo Flag$^+$ antibody responses in WBB mice versus mice in which the first dose and fate-mapping were omitted (ØBB). ØBB data are for ten mice from two independent experiments. WBB data are from **c**, remeasured in the same assay as ØBB. **g**, Schematic of antibody fractionation for the neutralization assay. **h**, BA.1 $NT_{50}$ titres for WBB mice at the same timepoint as in **d**. Post-depletion $NT_{50}$ values were normalized as described in the Methods. *P* values were calculated using one-tailed paired *t*-tests. **i**, The potency of the anti-RBD BA.1 Strep$^+$ (Flag-depleted) and Flag$^+$ (Strep-depleted) fractions. Raw $NT_{50}$ values for each fraction were divided by their respective RBD ELISA titres. *P* values were calculated using two-tailed paired *t*-tests. **j**, WH1 RBD structure (PDB: 6M0J) with amino acid changes in BA.1 highlighted in red. **k**, Deep mutational scanning analysis of serum samples obtained from three WBB mice 2 weeks after third immunization as in **d**. Antibody-binding sites on RBD are shaded according to escape fraction. The positions that are most highly targeted by each serum fraction are indicated (BA.1-specific residues are shown in parentheses).

To determine whether the enhanced neutralization of BA.1 observed after heterologous boosting (Fig. 4b) was due to the induction of new BA.1-specific antibodies in this setting, we fractionated WBB serum samples taken 2 weeks after the third immunization into Flag-depleted (Strep$^+$, primary) and Strep-depleted (Flag$^+$, new) preparations (Fig. 4g and Extended Data Fig. 5a) and measured their neutralizing potency against WH1 and BA.1 pseudovirus. As expected, depletion of Flag$^+$ antibodies had minimal effect on WH1 neutralization, whereas Strep$^+$ depletion resulted in

a much greater decrease (1.7-fold versus 8.5-fold, respectively; Extended Data Fig. 5b,c). By contrast, removing new ($Flag^+$) antibodies from WBB sera resulted in a greater reduction in BA.1 neutralization compared with removing primary ($Strep^+$) antibodies (4.9-fold versus 1.9-fold decrease; Fig. 4h), even though $Strep^+$ antibodies bound more avidly to the BA.1 RBD by ELISA (Fig. 4d). To estimate the neutralizing potency per unit of specific antibody, we divided the 50% neutralization titre ($NT_{50}$) derived from the BA.1 pseudovirus assay by the end-point binding titre obtained by BA.1 RBD-specific ELISA. When normalized to reactivity in this manner, new ($Flag^+$) antibody was on average 7.0-fold more potent at neutralizing BA.1 than the $Strep^+$ antibody produced by primary-cohort B cell clones in response to WH1 (Fig. 4i). As calculated in Extended Data Fig. 5d, around 80% of the excess BA.1 neutralization by WBB compared with WWW samples could be attributed to de novo generation of BA.1-specific antibodies from naive cells rather than to secondary affinity maturation or preferential selection of primary-cohort memory B cells. Thus, the antibodies that escape primary addiction after BA.1 boosting are optimized to neutralize the variant strain.

Finally, to gain mechanistic insights into how antigenic drift leads to attenuation of primary addiction, we performed deep mutational scanning[31,32] to define the WH1 and BA.1 RBD epitopes targeted by $Flag^+$ and $Strep^+$ antibodies in heterologously boosted mice (Fig. 4j and Extended Data Fig. 6). This approach enabled us to separately determine the epitopes targeted by primary and new antibody in the same mouse. We measured the antibody escape patterns of four mice against BA.1 (both $Strep^+$ and $Flag^+$ antibodies) and WH1 RBD ($Strep^+$ antibodies only), three of which showed interpretable dominance peaks (Extended Data Fig. 7). In these three mice, there was clear segregation of the epitopes targeted by primary $Strep^+$ and new $Flag^+$ antibodies (Fig. 4k and Extended Data Fig. 7). In two mice, $Strep^+$ antibodies targeted residues of the 'class 3' epitope located on the outer face of the RBD (Arg346, Arg357, Ile468), which, as expected, were conserved between WH1 and BA.1 (Fig. 4j,k). By contrast, $Flag^+$ antibodies in both mice were focused on the BA.1-specific residue Arg493 (Gln493 in WH1), located on the top of the RBD in the 'class 2' region of the ACE2-binding surface. The third mouse showed analogous segregation between

primary and new antibodies but targeted to different epitopes. Whereas Strep$^+$ antibodies were heavily focused on the conserved class 1/2 epitope that includes Gly485/Phe486 (on the top of the RBD at the ACE2 interface), Flag$^+$ antibodies bound primarily to an epitope that includes the BA.1-specific Lys440 residue (Asn440 in WH1) on the side face of the RBD distal to Gly485/Phe486 (class 3). Notably, both N440K and Q493R have been reported to lead to escape from neutralization by various monoclonal antibodies[33,34,35]. Thus, all three mice followed a logic in which new antibodies elicited by heterologous immunization preferentially targeted epitopes that contained BA.1-specific escape mutations and that did not overlap with epitopes bound by cross-reactive primary antibodies. We conclude that primary addiction, by acting in an epitope-specific manner, suppresses the de novo generation of antibodies to conserved epitopes, while allowing the induction of new antibodies targeted specifically to drifted epitopes.

# Discussion

Taken together, on the basis of our findings using the K-tag system, we make two main points. First, suppression of de novo antibody responses by existing immunity, a necessary feature of OAS as we define it, is extremely potent when measured at zero antigenic distance. These findings support the primary addiction/OAS[2] model (Fig. 2a) in which existing responses prevent the emergence of new serum antibodies to the same antigen, over a sequential contribution/seniority[20] model in which first-cohort responses are larger simply because they were established first and therefore boosted a greater number of times. Second, primary addiction weakens markedly as antigenic distance between priming and boosting strains increases. This observation suggests an explanation for why OAS has been so difficult to document experimentally in a consistent manner[5]—traditional measurements, which rely on differences between drifted antigens to assign antibodies to primary or de novo cohorts, may be able to distinguish these cohorts reliably only when antigenic distance is too great to allow for clear detection of primary addiction. A case in point is that our model estimates that primary addiction between PR8 and FM1, the influenza virus strains for which OAS was initially described[1,19], is relatively weak (equivalent to a

3.8-fold suppression of the de novo response), and may therefore be difficult to ascertain without the precision afforded by the K-tag model.

Our data are consistent with observations in humans showing that, after seasonal influenza vaccination, a large proportion of serum antibody clonotypes are recalled from pre-existing pools, although these studies remained agnostic to whether vaccine-induced antibody clonotypes arose from de novo responses or from recall of undetected memory B cells[17,36]. Humans, even with their rich influenza antigen exposure history, are able to mount substantial de novo responses in recall GCs, while relying on memory B cells for the immediate plasma cell response[37]. We therefore expect the general mechanics of primary addiction reported here to apply to humans as well, at least in general terms. In the SARS-CoV-2 setting, OAS-type suppression may explain the small and variable differences in the preferential induction of Omicron neutralization by variant-specific or bivalent compared with homologous vaccination[38,39,40,41,42]. Our data suggest that a second dose of an Omicron-containing vaccine may be required to reveal the full extent of de novo antibody induction by Omicron boosting. However, in practice, the repeated exposure of most individuals to WH1 antigens before Omicron boosting, as well as the potential greater propensity of human GCs to allow entry of reactivated memory B cells[37], may lead to stronger OAS-type suppression of Omicron-specific antibodies in real-life scenarios.

Mechanistically, our measurements at zero antigenic distance indicate a functional divide between recall GCs and serum responses in mice—whereas the former consist almost exclusively of naive-derived B cell clones[21,22,23], the latter are dominated by the effects of primary addiction. A potential explanation for this divergence is that the naive B cells that contribute to secondary GCs are not of sufficient affinity either to exit the GC as plasma cells or to secrete antibodies that are detectable by direct ELISA. Low antigen binding among naive-derived secondary GC B cells has been reported previously by others[22], which is consistent with the latter hypothesis. As antigenic drift for both influenza virus and SARS-CoV-2 is primarily driven by immune escape, the loosening of OAS in heterologous settings should in principle focus newly recruited B cell clones on drifted neutralizing epitopes. This view is supported by the results of our

neutralization and epitope mapping experiments. Moreover, RBD-binding antibodies that escaped primary addiction tended to focus on new residues located within epitopes that were not dominant targets of the cross-reactive first-cohort response. This pattern suggests antibody-mediated epitope masking—whereby serum antibodies that are either present before boosting or produced acutely by boosted memory B cells compete with naive B cells for binding to a specific epitope—as a potential mechanism for primary addiction. Similar effects have been observed previously by infusion of monoclonal antibodies before induction of an immune response in mice and humans and are predicted to affect the fine specificity of recall responses[43,44,45,46]. These findings not only indicate that the suppression of de novo antibody responses afforded by primary addiction is epitope specific rather than antigen specific (epitope masking, rather than antigen trapping[47]), but also suggest a teleological explanation for why it might be advantageous for memory B cells to avoid re-entering secondary GCs[21,48], as competition by memory B cells could inhibit the ability of naive cells to generate antibodies tailored to new epitopes in viral escape variants. In such a framework, the primary role of secondary GCs would be to circumvent the worst effects of OAS.

# Methods

## Mice

WT C57BL/6J and *B6.C(Cg)-Cd79a*$^{tm1(cre)Reth}$*/EhobJ* (*Cd79a*$^{Cre/+}$, also known as Mb1-Cre[49]) mice were obtained from The Jackson Laboratory. *S1pr2-cre*$^{ERT2}$ BAC-transgenic mice[25] were a gift from T. Kurosaki and T. Okada. *Igk*$^{Tag}$ mice were generated at the Rockefeller University. We designed the allele as indicated in Fig. 1a and Extended Data Fig. 1, with a Flag tag (DYKDDDDK) and Strep-II tag (WSHPQFEK) separated by stop codons and the SV40 poly-A transcriptional terminator. The 522-nucleotide single-stranded DNA template (including 5′ and 3′ homology arms, each 100 nucleotides long) and the CRISPR guide-RNA (GGAGCTGGTGGTGGCGTCTC) were purchased from IDT and prepared according to the Easi-CRISPR gene targeting method[50] by the Rockefeller University Gene Targeting Resource Center and injected into zygotes

obtained from C57BL/6 mice by the Rockefeller University Transgenic Services core facility. We verified that the allele was correctly inserted by Sanger sequencing across the entire locus using genomic primers located outside of the homology arms. To decrease the risk of potential CRISPR off-target effects, one founder mouse was back-crossed for at least five generations onto C57BL/6J mice before use in experiments. To generate $Igk^{Flag/Strep}$ mice, germline-excised (Cre-negative mice displaying $Strep^{+}$ B cell surface staining) mice, obtained from occasional spontaneous germline recombination in $Cd79a^{Cre/+}Igk^{Tag}$ breedings, were crossed to the parental $Igk^{Tag}$ strain. All of the mice were held at the immunocore clean facility at the Rockefeller University under specific-pathogen-free conditions. All mouse procedures were approved by the Rockefeller University's Institutional Animal Care and Use Committee.

## Immunizations, infections and treatments

Immune responses were induced in mice (male and female, aged 7–12 weeks) by either subcutaneous footpad immunization with 10 µg $TNP_{17}$ -KLH (Biosearch, T-5060) supplemented with 1/3 volume of Imject alum (Thermo Fisher Scientific) or i.p. immunization with 50 µg TNP-KLH prepared with 1/3 volume of either alum or aluminium hydroxide gel (alhydrogel, Invivogen); 20 µg recombinantly produced trimer-stabilized HA (see below) in alhydrogel; intramuscular immunization of quadricep muscles with 3 µg WH1 (ref. [51]) or BA.1 spike mRNA encapsulated in lipid nanoparticles (mRNA–LNP), generated as described below; or by intranasal infection with mouse-adapted PR8 influenza virus produced in embryonated chicken eggs (~33 plaque-forming units, provided by M. Carroll). In *S1pr2*-$Igk^{Tag}$ mice, the primary immune response was fate-mapped by oral gavage of 200 µl tamoxifen (Sigma-Aldrich) dissolved in corn oil at 50 mg $ml^{-1}$, on days 4 and 8 for the day 12 flow cytometry experiment (Fig. 1), on days 4, 8 and 12 for all other immunization experiments, and on days 4, 8, 12 and 16 for the influenza infection experiment (Fig. 3). In TNP recall experiments (Fig. 2c), boosting was performed identically to the primary immunization, at the indicated time points. For homologous mRNA–LNP experiments (Fig. 2e), boosting was performed in the same way as priming, except that, for some mice, the left quadricep muscle was boosted (contralaterally, stratified in Extended Data Fig. 3e). For heterologous mRNA–LNP recall experiments

(Fig. 4), boosting was contralateral in all cases. For HA immunization experiments (Fig. 3), homologous or heterologous HA protein boosting was performed ipsilaterally, exactly as the primary immunization. For infection experiments, mice were boosted after 3 months by subcutaneous immunization of the footpad with 5 μg $HA_{PR8}$ or $HA_{FM1}$ prepared with 1/3 volume alhydrogel, as described previously[21]. In all cases, additional booster immunizations were performed identically to the first boost. Blood samples were collected by cheek puncture into microtubes prepared with clotting activator serum gel (Sarstedt, 41.1378.005).

## Generation of recombinant and hapten-conjugated proteins

Recombinant HAs used for immunizations and ELISAs were produced in-house using the CHO cell protein expression system, as described previouosly[21]. Cysteine residues were introduced into the HA sequence to create trimer-stabilizing disulfide bonds, as originally described for H1/A/California/07/2009 (ref. [52]). We described the production of $HA_{PR8}$ and $HA_{CA09}$ previously[21]. For $HA_{FM1}$ and $HA_{NC99}$, the same procedure was followed, including the introduction of trimer-stabilizing mutations. For immunizations, C-terminal domains not native to HA (foldon, Avi-tag, His-tag) were removed by thrombin cleavage and HAs were subsequently fast protein liquid chromatography (FPLC)-purified before storage in phosphate-buffered saline (PBS). For ELISA, non-thrombin treated FPLC-purified proteins were used. A high-affinity IgY-specific monoclonal antibody obtained from CGG-immunized mice (clone 2.1 (ref. [53])) was modified for use as a standard for Flag/Strep ELISA detection. Heavy and light chain constant regions in the original human monoclonal antibody plasmids[54] were replaced with the mouse $IgG_1$ and Igκ constant regions and the C terminus of $C_\kappa$ was modified to encode a LoxP site and a Ser-Gly-Gly linker followed by either a Flag or Strep-tag, yielding $C_\kappa$ chains that were identical to those produced by *Igk*$^{Tag}$ mice before and after recombination, respectively. The mAb-Flag or mAb-Strep light-chain plasmids were transfected together with the heavy chain plasmid into HEK293F cells (Thermo Fisher Scientific, R79007) and purified using protein-G affinity chromatography as described previously[53]. Cell lines tested negative for mycoplasma and were not authenticated besides testing the protein they produced. To compare affinity

maturation between Flag$^+$ and Strep$^+$ anti-TNP antibody titres (Fig. 1h), custom low- and high-hapten bovine serum albumin (BSA) conjugations were made in-house. BSA (in PBS; Thermo Fisher Scientific, 77110) at 2.5 mg ml$^{-1}$ was incubated with TNP-ε-aminocaproyl-OSu (Biosearch Technologies, T-1030) in PBS with 20% dimethyl sulfoxide at a molar ratio of either 1:2 or 1:20 for 2 h at room temperature while rotating. Unconjugated TNP-ε-Aminocaproyl-OSu was removed by dialysis in PBS. Final TNP:BSA conjugation ratios were estimated to be ~1:1 and ~1:13 by measuring absorbance at 280 and 348 nm, these reagents are referred to as $TNP_1$-BSA and $TNP_{13}$-BSA. The BSA concentration was corrected by determining the extinction coefficient for TNP-ε-aminocaproyl-OSu at 280 nm. Except for in Fig. 1h, commercial $TNP_4$-BSA (Biosearch Technologies, T-5050) was used for all other TNP ELISAs (described below).

## Production of mRNA–LNP

The WH1 S mRNA vaccine was designed on the basis of the SARS-CoV-2 S protein sequence (Wuhan-Hu-1, GenBank: MN908947.3) where the lysine and valine amino acids in positions 986–987 were modified to proline residues to obtain a prefusion-stabilized mRNA-encoded immunogen. The BA.1 S amino acid sequence was obtained from the WH1 S by introducing BA.1-specific modifications. Coding sequences of the WH1 and BA.1 S were codon-optimized, synthesized and cloned into an mRNA production plasmid (GenScript) as described previously[55]. mRNA production and LNP encapsulation was performed as described previously[55]. In brief, mRNAs were transcribed to contain 101 nucleotide-long poly(A) tails. m1Ψ-5′-triphosphate (TriLink) instead of UTP was used to generate modified nucleoside-containing mRNAs. Capping of the in vitro transcribed mRNAs was performed co-transcriptionally using the trinucleotide cap1 analogue, CleanCap (TriLink). mRNA was purified by cellulose (Sigma-Aldrich) purification, as described previously[56]. All mRNAs were analysed by agarose gel electrophoresis and were stored frozen at −20 °C. Cellulose-purified m1Ψ-containing RNAs were encapsulated in LNPs using a self-assembly process as previously described, whereby an ethanolic lipid mixture of ionizable cationic lipid, phosphatidylcholine, cholesterol and polyethylene glycol-lipid was rapidly mixed with an aqueous solution

containing mRNA at acidic pH[57]. The ionizable cationic lipid and LNP composition are described in the patent application WO 2017/004143. The RNA-loaded particles were characterized and subsequently stored at −80 °C at a concentration of 1 µg $\mu l^{-1}$. The mean hydrodynamic diameter of these mRNA–LNPs was around 80 nm with a polydispersity index of 0.02–0.06 and an encapsulation efficiency of approximately 95%.

## Flow cytometry

For flow cytometry analysis of peripheral B cells, blood was collected in microtubes with ethylenediaminetetraacetic acid (EDTA) to prevent coagulation and treated with ammonium–chloride–potassium (ACK) buffer (Lonza) to lyse red blood cells. For lymph node samples, cell suspensions were obtained by mechanical disassociation with disposable micropestles (Axygen). Spleens were homogenized by filtering through a 70 µm cell strainer and treated with ACK buffer. Bone marrow cells were extracted by centrifugation of punctured tibiae and femurs at up to 10,000*g* for 10 s, then treated with ACK buffer. Cells from each tissue were resuspended in PBS supplemented with 0.5% BSA and 1 mM EDTA and incubated first with FC-block (rat anti-mouse CD16/32, 2.4G2, Bio X Cell) for 30 min on ice and subsequently with various fluorescently labelled antibodies (Supplementary Table 1) for 30 min. Cells were filtered and washed with the same buffer before analysis on a BD FACS Symphony cytometer. Data were analysed using FlowJo (v.10).

## Western blotting

To determine the presence of epitope-tagged antibodies in $Igk^{Tag}$ mice, serum samples from steady state adult mice and precision plus dual-colour protein standards (Bio-Rad) were run in triplicate on SDS–PAGE mini-protean TGX protein gels (Bio-Rad) under denaturing conditions, to separate heavy and light antibody chains. The samples were transferred to a polyvinylidene difluoride membrane using the Iblot gel transfer system (Invitrogen). Membranes were blocked for 2 h at room temperature while gently shaking with 5% non-fat dry milk in PBS-Tween-20 (0.05%), before overnight incubation in the same buffer with 1:2,000 anti-Flag-HRP (D6W5B, Cell Signalling Technology, 86861S) or anti-Strep (Strep-tag II

StrepMAB-Classic, Bio-Rad, MCA2489P) or goat anti-mouse Igκ-HRP (Southern Biotech, 1050-05). Membranes were extensively washed with PBS-Tween-20 and subsequently incubated with western blotting ECL substrate (Amersham) before chemiluminescence detection using the Azure c300 gel imager (Azure Biosystems).

## ELISA

ELISAs were performed as described previously[21], with specific modifications to allow for direct Flag/Strep comparison. Flag/Strep ELISAs were performed side by side and with internal standards on each 96-well plate. To detect antigen-specific serum antibody titres, the plates were coated overnight at 4 °C with antigen in PBS (10 μg ml$^{-1}$ for $TNP_4$-BSA, 2 μg ml$^{-1}$ for in-house-conjugated $TNP_{1/13}$-BSA (see above), and 1 μg ml$^{-1}$ for HAs and SARS-CoV-2 spike or RBD proteins (Sinobiological, 40592-V08H, 40592-V08H121, 40589-V08H26; WH1 S and RBD proteins were a gift from P. Wilson)). For Flag/Strep standard curves, wells were coated with 10 μg ml$^{-1}$ purified IgY (Gallus Immunotech). After washing with PBS-Tween (PBS + 0.05% Tween-20, Sigma-Aldrich), plates were blocked for 2 h at room temperature with 2.5% BSA in PBS. Serum samples were diluted 1:100 in PBS and serially titrated in threefold dilutions. Mouse anti-IgY mAb-Flag or mAb-Strep were also serially titrated in threefold dilutions (Extended Data Fig. 2d). The samples were incubated for 2 h and then washed with PBS-Tween, before adding one of the following HRP-detection antibodies: goat anti-mouse IgG (Jackson Immunoresearch, 15-035-071), rat anti-mouse Igκ (Abcam, ab99632), goat anti-mouse IgM (Southern Biotech, 1020-05), rabbit anti-Flag-HRP (D6W5B) or mouse anti-Strep (Strep-tag II StrepMAB-Classic) for 30–45 min. Dilutions of anti-Flag and anti-Strep antibodies were defined so that the curves generated by titration of Flag- and Strep-tagged monoclonal antibodies were equivalent (Extended Data Fig. 2d). After washing with PBS-Tween, samples were incubated with 3,3′,5,5′-tetramethylbenzidine substrate (slow kinetic form, Sigma-Aldrich) and the reaction was stopped with 1 N HCl. Optical density (OD) absorbance was measured at 450 nm on the Fisher Scientific accuSkan FC plate reader. To normalize Flag and Strep end-point titres, the serum titre dilution was calculated at which each sample passed the threshold OD value of its respective monoclonal antibody at a fixed concentration of either 20 or 6.67

ng μl$^{-1}$. Titres were calculated by logarithmic interpolation of the dilutions with readings immediately above and immediately below the monoclonal antibody OD used[58].

For total serum IgG ELISAs, plates were coated with anti-mouse IgG. Standard curves were generated using unlabelled mouse IgG (Southern Biotech, 0107-01), and detection was performed using anti-mouse IgG-HRP (Southern Biotech). To deplete IgM from the serum samples, anti-mouse IgM agarose beads (Sigma-Aldrich, A4540) were used according to the manufacturer's instructions. Beads were washed with PBS and samples were incubated at a ratio of 1:20 sample to beads overnight at 4 °C with rotation. The bead-bound IgM fraction was removed by centrifugation for 3 min at 10,000*g*, and the unbound supernatant fraction was used for subsequent ELISAs. To confirm the efficiency of IgM depletion, total IgM levels were measured as described above for total IgG, with goat anti-mouse IgM, unlabelled IgM and anti-mouse IgM-HRP (Southern Biotech).

## Serum fractionation and SARS-CoV-2 pseudoneutralization assay

Virus neutralization titres were assessed in the Flag$^{+}$ versus Strep$^{+}$ serum fractions of samples collected from S mRNA–LNP immunized *S1pr2-Igk$^{Tag}$* mice. To separate fractions, immunoprecipitation with anti-Flag M2 magnetic beads (Sigma-Aldrich, M8823-5ML) and MagStrep type 3 XT beads (IBA, 2-4090-010) was performed according to the manufacturer's instructions. In brief, magnetic beads were washed with sample buffer (Tris-buffered saline for Flag beads, and 1× buffer W (IBA, 2-1003-100) for MagStrep beads) and the samples were incubated at a ratio of 20:1 sample to bead resin overnight at 4 °C with rotation. Bead-bound fractions were separated using a magnetic separator and discarded, while the unbound fraction was collected. Fractionated samples were concentrated by centrifugation to half the input concentration and heat inactivated. The degree to which the total and fractionated serum samples neutralized WH1 and BA.1 SARS-CoV-2 was approximated using SARS-CoV-2 spike pseudotyped HIV-1 based NanoLuc luciferase reporter assay described previously[29]. In brief, serum samples were five-fold serially diluted with a final top dilution of 1:00 serum and incubated for 1 h at 37 °C with SARS-

CoV-2 WH1 or BA.1 spike pseudotyped HIV-1 reporter virus and then transferred to HT1080/ACE2.cl14 cells[59]. At 48 h, the cells were washed and lysed and the luciferase activity was measured using the Nano-Glo Luciferase Assay System (Promega) and the Glomax Navigator luminometer (Promega). The relative luminescence units were normalized using cells infected in the absence of serum and then plotted in GraphPad Prism. $NT_{50}$ values were calculated using four-parameter nonlinear regression (least squares regression method without weighting) of the curves shown in Extended Data Fig. 5b. The mean of two technical duplicates is shown, outlier points were excluded. For $NT_{50}$ comparisons between the input and fractions (Fig. 4g and Extended Data Fig. 5c), the $NT_{50}$ of the fractionated samples was adjusted to equalize the BA.1 RBD ELISA titre of the undepleted tag compared to its corresponding ELISA titre in the input fraction.

## Deep mutational scanning

### Construction of yeast-displayed deep mutational scanning libraries of Omicron BA.1 RBD

Duplicate single-mutant site-saturation variant libraries were designed in the background of the SARS-CoV-2 Omicron BA.1 spike RBD and produced by Twist Bioscience, essentially the same as has been done previously for other SARS-CoV-2 variants[60,61]. The GenBank map of the plasmid encoding the unmutated Omicron BA.1 RBD in the yeast-display vector is available at GitHub (https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS_Omicron/blob/main/data/3294_pETcon-SARS2-RBD_Omicron-BA1.gb). The site-saturation variant libraries were delivered as double-stranded DNA fragments by Twist Bioscience and were barcoded and cloned in bulk into the yeast-display vector backbone. The barcoded mutant library plasmid DNA was electroporated into *Escherichia coli* (NEB 10-beta electrocompetent cells, New England BioLabs, C3020K), and bottlenecked to around $1 \times 10^5$ CFU (an average of >25 barcodes per single mutant). Plasmid DNA was purified and transformed into the AWY101 yeast strain. Sixteen-nucleotide barcodes were associated with their BA.1 variants by PacBio sequencing, and the effects of mutations of RBD

expression and ACE2 binding were measured, essentially as described[61]. These experiments are described and analysed at GitHub (https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS_Omicron).

### FACS sorting of yeast libraries to select mutations with reduced binding by polyclonal sera from immunized mice

Experiments mapping mutations that reduce RBD binding of sera from immunized mice were performed in biological duplicate with independent mutant WH1 or BA.1 RBD libraries, similarly to as previously described for monoclonal antibodies[62] and human polyclonal plasma samples[63]. First, 75 µl of each of the sera was twice-depleted of non-specific yeast-binding antibodies by incubating for 2 h at room temperature or overnight at 4 °C with 37.5 OD units of AWY101 yeast containing an empty vector, as described previously[63]. WH1 and BA.1 mutant RBD yeast libraries[61] were induced with galactose-containing, low-dextrose synthetic defined medium with casamino acids (SD-CAA, 6.7 g $l^{-1}$ yeast nitrogen base, 5.0 g $l^{-1}$ casamino acids, 1.065 g $l^{-1}$ MES acid and 2% (w/v) galactose + 0.1% (w/v) dextrose) to express RBD, then washed and incubated with diluted serum for 1 h at room temperature with gentle agitation. Each tested combination of mouse serum against each WH1 or BA.1 RBD mutant library for loss of binding of Strep or Flag-tag antibodies was performed independently. For each serum, a subsaturating dilution was used such that the amount of fluorescent signal due to serum antibody binding to RBD was approximately equal across samples (1:1,000 for mapping of Strep antibodies against the WH1 libraries; 1:200 for mapping of Strep antibodies against the BA.1 libraries; and 1:50 for mapping of the Flag antibodies against the BA.1 libraries). The yeast libraries were then secondarily labelled for 1 h with 1:100 FITC-conjugated anti-MYC antibodies (Immunology Consultants Lab, CYMC-45F) to label for RBD expression and either 1:200 APC-conjugated Streptavidin (Invitrogen S-868) to label for bound Strep antibodies or APC-conjugated rat anti-Flag (BioLegend, 637308) to label for bound Flag-tagged antibodies. A flow cytometry selection gate was drawn to capture RBD mutants with reduced antibody binding for their degree of RBD expression. For each sample, around $4 \times 10^6$ cells were processed on the BD FACSAria II cell sorter. Serum-escaped cells were grown overnight in SD-CAA as defined above with 2% (w/v) dextrose, no galactose, and 100

U ml$^{-1}$ penicillin + 100 µg ml$^{-1}$ streptomycin to expand cells before plasmid extraction.

### DNA extraction and Illumina sequencing

Plasmid DNA was extracted from 30 OD units ($1.6 \times 10^8$ colony-forming units (CFU)) of preselection yeast populations and approximately 5 OD units (around $3.2 \times 10^7$ CFU) of overnight cultures of serum-escaped cells (Zymoprep Yeast Plasmid Miniprep II) as previously described[60,62]. The 16-nucleotide barcodes identifying each WH1 or BA.1 RBD variant were amplified by polymerase chain reaction and prepared for Illumina sequencing as described previously[60,62]. Barcodes were sequenced on the Illumina NextSeq 2000 system with 50 bp single-end reads.

### Analysis of deep sequencing data to compute each mutation's escape fraction

Escape fractions were computed essentially as described previously[62]. We used the dms_variants package (https://jbloomlab.github.io/dms_variants/, v.1.4.0) to count each barcoded RBD variant in each preselection and serum-escape population. For each selection, we computed the escape fraction for each barcoded variant by the formula provided in ref. [62]. These escape fractions represent the estimated fraction of cells expressing that specific variant that falls in the escape bin, such that a value of 0 means the variant is always bound by serum and a value of 1 means that it always escapes serum binding. We then applied a computational filter to remove variants with >1 amino-acid mutation, low sequencing counts (<50 in the preselection condition) or highly deleterious mutations that might cause antibody escape simply by leading to poor expression of properly folded RBD on the yeast cell surface (an ACE2 binding score of <−2 or an RBD expression score of <−1.25 or <−0.83361 for the WH1 and BA.1 mutant libraries, respectively, reflecting the different baseline expression levels of the two wild-type RBDs). The reported antibody-escape scores throughout the paper are the average across duplicate libraries; these scores are also in Supplementary Table 1. Correlations in final single-mutant escape scores are shown in Extended Data Fig. 6c. Full documentation of the computational analysis is

available at GitHub (https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_OAS).

### Data visualization

The serum-escape map logo and line plots were created using the dmslogo package (https://jbloomlab.github.io/dmslogo, v.0.6.2). The height of each letter indicates the escape fraction for that amino-acid mutation. For each serum, the logo plots feature any site where for ≥1 library/antibody tag condition, the site-total antibody escape was >10× the median across all sites and at least 10% the maximum of any site. For each sample, the *y* axis was scaled to be the greatest of (1) the maximum site-wise escape metric observed for that sample or (2) 20× the median site-wise escape fraction observed across all sites for that plasma. The code that generates these logo plot visualizations is available at GitHub (https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_OAS/blob/main/results/summary/escape_profiles.md). To visualize serum escape on the RBD structure, the WH1 RBD surface (PDB: 6M0J) was coloured by the site-wise escape metric at each site, with white indicating no escape and red indicating the site with the most escape.

## Statistical analysis and software

Statistical tests used to compare conditions are indicated in figure legends. No statistical methods were used to determine the sample size. Statistical analysis was carried out using GraphPad Prism v.9. Flow cytometry analysis was carried out using FlowJo (v.10). Graphs were plotted using Prism (v.9) and edited for appearance using Adobe Illustrator CS. For data plotted on logarithmic scales (such as serum antibody titres), statistical analysis was performed on the log-transformed data. Samples with reactivities below the limit of detection were assigned a value of 100, as the top dilution was 1:100.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

# Data availability

The raw Illumina reads of the 16-nucleotide variant barcodes from the deep mutational scanning experiments are available at the NCBI SRA under BioProject PRJNA770094, BioSample SAMN30086726. All escape scores are shown in Supplementary Table 1 and are available at GitHub (https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_OAS/blob/main/results/supp_data/all_raw_data.csv). Renderings of spike RBD and HA structures were obtained from the PDB under accession codes 6M0J, 1RU7 and 3LZG.

# Code availability

The full code used to analyse the deep mutational scanning experiments is available at GitHub (https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_OAS). The code to generate the logo plot visualizations is available at GitHub (https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_OAS/blob/main/results/summary/escape_profiles.md).

# References


1. Francis, T. On the doctrine of original antigenic sin. *Proc. Am. Phil. Soc.* **104**, 572–578 (1960).

2. Fazekas de St Groth, S. & Webster, R. G. Disquisitions on original antigenic sin. II. Proof in lower creatures. *J. Exp. Med.* **124**, 347–361 (1966).

3. Cobey, S. & Hensley, S. E. Immune history and influenza virus susceptibility. *Curr. Opin. Virol.* **22**, 105–111 (2017).

4. Henry, C., Palm, A. E., Krammer, F. & Wilson, P. C. From original antigenic sin to the universal influenza virus vaccine. *Trends Immunol.* **39**, 70–79 (2018).

5. Yewdell, J. W. & Santos, J. J. S. Original antigenic sin: how original? How sinful? *Cold Spring Harb. Perspect. Med.* **11**, a038786 (2021).

6. Krammer, F. The human antibody response to influenza A virus infection and vaccination. *Nat. Rev. Immunol.* **19**, 383–397 (2019).

7. Wheatley, A. K. et al. Immune imprinting and SARS-CoV-2 vaccine design. *Trends Immunol.* **42**, 956–959 (2021).

8. Monto, A. S., Malosh, R. E., Petrie, J. G. & Martin, E. T. The doctrine of original antigenic sin: separating good from evil. *J. Infect. Dis.* **215**, 1782–1788 (2017).

9. Victora, G. D. & Nussenzweig, M. C. Germinal centers. *Annu. Rev. Immunol.* **40**, 413–442 (2022).

10. Baumgarth, N. The shaping of a B cell pool maximally responsive to infections. *Annu. Rev. Immunol.* **39**, 103–129 (2021).

11. Bhattacharya, D. Instructing durable humoral immunity for COVID-19 and other vaccinable diseases. *Immunity* **55**, 945–964 (2022).

12. Pulendran, B. & Ahmed, R. Immunological mechanisms of vaccination. *Nat. Immunol.* **12**, 509–517 (2011).

13. Gaebler, C. et al. Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**, 639–644 (2021).

14. Tan, J. et al. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature* **529**, 105–109 (2016).

15. Davis, C. W. et al. Longitudinal analysis of the human B cell response to ebola virus infection. *Cell* **177**, 1566–1582 (2019).

16. Turner, J. S. et al. SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. *Nature* **596**, 109–113 (2021).

17. Lee, J. et al. Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nat. Med.* **22**, 1456–1464 (2016).

18. Han, J. et al. Polyclonal epitope mapping reveals temporal dynamics and diversity of human antibody responses to H5N1 vaccination. *Cell Rep.* **34**, 108682 (2021).

19. Francis, T. Jr. Influenza: the new acquayantance. *Ann. Intern. Med.* **39**, 203–221 (1953).

20. Lessler, J. et al. Evidence for antigenic seniority in influenza A (H3N2) antibody responses in southern China. *PLoS Pathog.* **8**, e1002802 (2012).

21. Mesin, L. et al. Restricted clonality and limited germinal center reentry characterize memory B cell reactivation by boosting. *Cell* **180**, 92–106 (2020).

22. Viant, C. et al. Antibody affinity shapes the choice between memory and germinal center B cell fates. *Cell* **183**, 1298–1311 (2020).

23. Kuraoka, M. et al. Recall of B cell memory depends on relative locations of prime and boost immunization. *Sci. Immunol.* **7**, eabn5311 (2022).

24. Mostoslavsky, R., Alt, F. W. & Rajewsky, K. The lingering enigma of the allelic exclusion mechanism. *Cell* **118**, 539–544 (2004).

25. Shinnakasu, R. et al. Regulated selection of germinal-center cells into the memory B cell compartment. *Nat. Immunol.* **17**, 861–869 (2016).

26. Smith, D. J., Forrest, S., Ackley, D. H. & Perelson, A. S. Variable efficacy of repeated annual influenza vaccination. *Proc. Natl Acad. Sci. USA* **96**, 14001–14006 (1999).

27. Fonville, J. M. et al. Antibody landscapes after influenza virus infection or vaccination. *Science* **346**, 996–1000 (2014).

28. Hogan, M. J. & Pardi, N. mRNA vaccines in the COVID-19 pandemic and beyond. *Annu. Rev. Med.* **73**, 17–39 (2022).

29. Schmidt, F. et al. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J. Exp. Med.* **217**, e20201181 (2020).

30. Muecksch, F. et al. Affinity maturation of SARS-CoV-2 neutralizing antibodies confers potency, breadth, and resilience to viral escape mutations. *Immunity* **54**, 1853–1868.e7 (2021).

31. Greaney, A. J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476 (2021).

32. Starr, T. N. et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850–854 (2021).

33. Liu, L. et al. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* **602**, 676–681 (2022).

34. McCallum, M. et al. Structural basis of SARS-CoV-2 Omicron immune evasion and receptor engagement. *Science* **375**, 864–868 (2022).

35. Weisblum, Y. et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* **9**, e61312 (2020).

36. Lee, J. et al. Persistent antibody clonotypes dominate the serum response to influenza over multiple years and repeated vaccinations. *Cell Host Microbe* **25**, 367–376 (2019).

37. Turner, J. S. et al. Human germinal centres engage memory and naive B cells after influenza vaccination. *Nature* **586**, 127–132 (2020).

38. Alsoussi, W. B. et al. SARS-CoV-2 Omicron boosting induces de novo B cell response in humans. Preprint at *bioRxiv* https://doi.org/10.1101/2022.09.22.509040 (2022).

39. Collier, A. Y. et al. Immunogenicity of BA.5 bivalent mRNA vaccine boosters. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMc2213948 (2023).

40. Wang, Q. et al. Antibody response to Omicron BA.4–BA.5 bivalent booster. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMc2213907 (2023).

41. Davis-Gardner, M. E. et al. Neutralization against BA.2.75.2, BQ.1.1, and XBB from mRNA bivalent booster. *N. Engl. J. Med.* **388**, 183–185 (2023).

42. Kurhade, C. et al. Low neutralization of SARS-CoV-2 Omicron BA.2.75.2, BQ.1.1 and XBB.1 by parental mRNA vaccine or a BA.5 bivalent booster. *Nat. Med.* https://doi.org/10.1038/s41591-022-02162-x (2022).

43. McNamara, H. A. et al. Antibody feedback limits the expansion of B cell responses to malaria vaccination but drives diversification of the humoral response. *Cell Host Microbe* **28**, 572–585 (2020).

44. Meyer-Hermann, M. Injection of antibodies against immunodominant epitopes tunes germinal centers to generate broadly neutralizing antibodies. *Cell Rep.* **29**, 1066–1073 (2019).

45. Tas, J. M. J. et al. Antibodies from primary humoral responses modulate recruitment of naive B cells during secondary responses. *Immunity* **55**, 1856–1871 (2022).

46. Schaefer-Babajew, D. et al. Antibody feedback regulates immune memory after SARS-CoV-2 mRNA vaccination. *Nature* **613**, 735–742 (2023).

47. Kok, A., Fouchier, R. A. M. & Richard, M. Cross-reactivity conferred by homologous and heterologous prime-boost A/H5 influenza vaccination strategies in humans: a literature review. *Vaccines* **9**, 1465 (2021).

48. Pape, K. A., Taylor, J. J., Maul, R. W., Gearhart, P. J. & Jenkins, M. K. Different B cell populations mediate early and late memory during an endogenous immune response. *Science* **331**, 1203–1207 (2011).

49. Hobeika, E. et al. Testing gene function early in the B cell lineage in mb1-cre mice. *Proc. Natl Acad. Sci. USA* **103**, 13789–13794 (2006).

50. Quadros, R. M. et al. Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biol.* **18**, 92 (2017).

51. Laczko, D. et al. A single immunization with nucleoside-modified mRNA vaccines elicits strong cellular and humoral immune responses against SARS-CoV-2 in mice. *Immunity* **53**, 724–732 (2020).

52. Lee, P. S., Zhu, X., Yu, W. & Wilson, I. A. Design and structure of an engineered disulfide-stabilized influenza virus hemagglutinin trimer. *J. Virol.* **89**, 7417–7420 (2015).

53. Tas, J. M. et al. Visualizing antibody affinity maturation in germinal centers. *Science* **351**, 1048–1054 (2016).

54. Wardemann, H. et al. Predominant autoantibody production by early human B cell precursors. *Science* **301**, 1374–1377 (2003).

55. Freyn, A. W. et al. A multi-targeting, nucleoside-modified mRNA influenza virus vaccine provides broad protection in mice. *Mol. Ther.* **28**, 1569–1584 (2020).

56. Baiersdorfer, M. et al. A facile method for the removal of dsRNA contaminant from in vitro-transcribed mRNA. *Mol. Ther. Nucleic Acids* **15**, 26–35 (2019).

57. Maier, M. A. et al. Biodegradable lipids enabling rapidly eliminated lipid nanoparticles for systemic delivery of RNAi therapeutics. *Mol. Ther.* **21**, 1570–1578 (2013).

58. Ersching, J. et al. Germinal center selection and affinity maturation require dynamic regulation of mTORC1 kinase. *Immunity* **46**, 1045–1058 (2017).

59. Schmidt, F. et al. Plasma neutralization of the SARS-CoV-2 Omicron variant. *N. Engl. J. Med.* **386**, 599–601 (2022).

60. Starr, T. N. et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310 (2020).

61. Starr, T. N. et al. Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science* **377**, 420–424 (2022).

62. Greaney, A. J. et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **29**, 44–57 (2021).

63. Greaney, A. J. et al. A SARS-CoV-2 variant elicits an antibody response with a shifted immunodominance hierarchy. *PLoS Pathog.* **18**, e1010248 (2022).

## Acknowledgements

We thank the staff at the Rockefeller University Transgenics and Gene Targeting facilities for generating the K-tag mouse strain and Comparative Biosciences Center for mouse housing; P. Wilson for recombinant SARS-CoV-2 WH1 spike and RBD protein; J. T. Jacobsen for technical assistance; and all of the Rockefeller University staff for their continuous support. This study was funded by NIH/NIAID grants R01AI119006 and R01AI139117 to G.D.V., P01AI165075 to P.D.B., R01AI146101 and R01AI153064 to N.P. and R01AI141707 to J.D.B. Work in the Victora laboratory is additionally supported by NIH grant DP1AI144248 (Pioneer award) and the Robertson Foundation. A.S. was supported by a Boehringer–Ingelheim Fonds PhD fellowship. P.D.B. and J.D.B. are HHMI investigators. G.D.V. is a Burroughs–Wellcome Investigator in the Pathogenesis of Infectious Disease, a Pew-Stewart Scholar and a MacArthur Fellow.

## Author information

## Authors and Affiliations

1. Laboratory of Lymphocyte Dynamics, The Rockefeller University, New York, NY, USA

   Ariën Schiepers, Marije F. L. van ’t Wout, Luka Mesin & Gabriel D. Victora

2. Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

   Allison J. Greaney, Tyler N. Starr & Jesse D. Bloom

3. Laboratory of Retrovirology, The Rockefeller University, New York, NY, USA

   Trinity Zang & Paul D. Bieniasz

4. Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

   Hiromi Muramatsu & Norbert Pardi

5. Acuitas Therapeutics, Vancouver, British Columbia, Canada

   Paulo J. C. Lin & Ying K. Tam

6. Howard Hughes Medical Institute, Chevy Chase, MD, USA

   Paul D. Bieniasz & Jesse D. Bloom

## Contributions

A.S. and M.F.L.v.’t.W. performed all experimental work, with essential input from L.M. A.J.G. performed and interpreted the deep mutational scanning experiments under the supervision of T.N.S. and J.D.B. N.P. and H.M. designed and produced WH1 and BA.1 S-encoding mRNAs. P.J.C.L. and Y.K.T. formulated mRNAs into LNP. T.Z. carried out pseudovirus-neutralization assays under the supervision of P.D.B. A.S. and G.D.V.

designed the K-tag allele and all of the experiments in the study and wrote the manuscript with input from all of the authors.

### Corresponding author

Correspondence to Gabriel D. Victora.

## Ethics declarations

### Competing interests

N.P. is named on a patent (62/153,143 — Nucleoside-modified RNA inducing an adaptive immune response), describing the use of nucleoside-modified mRNA in LNPs as a vaccine platform. He has disclosed those interests fully to the University of Pennsylvania and has an approved plan in place for managing any potential conflicts arising from the licensing of that patent. P.J.C.L. and Y.K.T. are employees of Acuitas Therapeutics, a company that is involved in the development of mRNA–LNP therapeutics. Y.K.T. is named on a patent (WO 2017/004143 — Lipids and lipid nanoparticle formulations for delivery of nucleic acids) that describes LNPs for the delivery of nucleic acid therapeutics, including mRNA, and the use of modified mRNA in lipid nanoparticles as a vaccine platform. P.D.B. has done consulting work in the area of COVID vaccines for Pfizer. J.D.B. consults or has recently consulted for Apriori Bio, Oncorus, Merck and Moderna on topics related to viruses, vaccines and viral evolution. J.D.B, T.N.S. and A.J.G. are listed as inventors on Fred Hutch licensed patents (62/935,954 and 62/692,398) related to viral deep mutational scanning. G.D.V. and J.D.B. are advisors for the Vaccine Company.

## Peer review

### Peer review information

*Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Design of the K-tag allele.

Nucleotide sequence of the 522 bp DNA template used to generate the *Igk*$^{Tag}$ allele, with amino acid translations given for all coding sequences (bold font). Nucleotide numbers are given for each line. Amino acid translation is positioned below the centre nucleotide of each codon.

### Extended Data Fig. 2 Characterization of the K-tag system.

(**a**) Representative flow cytometry plots of peripheral B cells (gated: B220$^{+}$CD4$^{-}$CD8$^{-}$CD138$^{-}$) obtained from the blood of WT and *Igk*$^{Tag/Tag}$ mice, stained for surface expression of Igλ and FLAG-tagged immunoglobulins. (**b**) Total IgG concentration in the serum of WT, *Igk*$^{Tag/Tag}$, and *Cd79a*$^{Cre/+}$. *Igk*$^{Tag/Tag}$ mice as determined by ELISA. Differences were not significant by one-way ANOVA ($p < 0.05$). (**c**) Flow cytometry of steady-state bone marrow plasma cells (PCs) obtained from adult (6-week old) *Igk*$^{FLAG/Strep}$ mice. Gating strategy for PCs is shown in the left panel (pre-gated on non-CD4/CD8 T cells), for surface IgA$^{+}$ and Igλ$^{-}$ PCs in the middle panel, and for FLAG/Strep in the right panel. Quantification across 9 mice from 3 independent experiments is shown in the rightmost panel (each dot represents an individual *Igk*$^{FLAG/Strep}$ mouse, line represents the median). (**d**) ELISA standard curves with monoclonal antibody (mAb)-FLAG and mAb-Strep detected at dilutions of the respective HRP antibodies where the curves overlap. The mAb concentration at which the curves crossed the absorbance background threshold (indicated by the dotted lines) was used to calculate the endpoint titre, as described in the methods section. (**e**) Tag-specific anti-TNP titres in *Igk*$^{FLAG/Strep}$ mice immunized and boosted i.p. with TNP-KLH/alum at the

timepoints indicated by black arrows. Results are from 14 mice from 2 independent experiments. The day 7 timepoint was not collected for the first cohort. Thin lines represent individual mice, thick lines link medians of log transformed titre values at each time point. (**f**) Flow cytometry of S1pr2-*Igk*$^{WT/Tag}$ mice as in Fig. 1d,e. with quantification in (**g**), cre$^{-}$ and no tamoxifen control groups are included. Data points are from 6—13 popliteal lymph nodes per group from at least 2 independent experiments. (**h**) Anti-TNP ELISA reactivity for S1pr2-*Igk*$^{Tag/Tag}$ mice immunized as in Fig. 1d. Serum was obtained at 47 d.p.i from 6 mice that received tamoxifen and 4 that did not receive tamoxifen. IgG (left), FLAG (middle) and Strep (right) ELISA absorbance is shown. Samples were diluted 1:100.

## Extended Data Fig. 3 Recall antibodies and GCs.

(**a**) Flow cytometry of secondary GCs in the spleen of 3 TNP-KLH i.p. primed and boosted S1pr2-*Igk*$^{Tag/Tag}$ mice, with tamoxifen labelling at 4, 8, and 12 d.p.i.. Gated on GC B cells (FAS$^{+}$CD38$^{-}$B220$^{+}$CD4$^{-}$CD8$^{-}$CD138$^{-}$) expressing FLAG or Strep-tag as in Fig. 1e. (**b**) Total IgM concentrations of serum samples pre and post IgM-depletion via immunoprecipitation, measured by ELISA. Data are from 8 serum samples, from the same 4 mice as in Fig. 2c. (**c**) Tag-specific anti-TNP titres before and after IgM depletion for samples collected 6 days after the first and second boost with TNP-KLH (same samples as in (b) and Fig. 2c). Bars represent the means of log transformed titres and the error bars are SEM. P-values are for two-tailed, paired T-test, only statistically significant ($p < 0.05$) values are shown. (**d**) Background Strep$^{+}$ anti-RBD titres in S1pr2-*Igk*$^{Tag/Tag}$ control mice immunized as in Fig. 2b,e, but not treated with tamoxifen. Graphs show median percentage of the anti-RBD titre that is Strep$^{+}$ ((S/(S + F)*100)) in the absence of tamoxifen at the pre-boost time point and two weeks after the second and third immunizations. This represents the median percentage by which FLAG$^{+}$ titres in recall responses are likely to be underestimated by spontaneous recombination by the S1pr2-CreERT2 driver. Data are from 8 mice from 2 independent experiments. (**e**) Comparison of primary addiction data shown in Fig. 2e,f, stratified by cohort and ipsilateral versus contralateral boost. The 4$^{th}$ cohort of mice is the homologously boosted group shown in Fig. 4b—e, depicted here by open circles. Bars represent the mean of log transformed titres, error bars are SEM. Data are from 16 mice

from 4 independent cohorts. (**f**) Comparison of primary addiction between 3rd and 4th responses in 9 mice from 2 cohorts, based on data from Fig. 2e,f,h. Bars represent the mean of log transformed titres, error bars are SEM.

## Extended Data Fig. 4 Primary cohort recall upon heterologous HA protein boosting.

(**a**) Comparison of *de novo* $FLAG^+$ antibody responses to $HA_{FM1}$ in the presence or absence of primary infection with influenza virus PR8. PR8>FM1>FM1 data are for the same samples as in Fig. 3d, re-measured in the same assay as Ø>FM1>FM1. (**b**) Schematic representation of HA prime-boost strategy. S1pr2-$Igk^{Tag/Tag}$ mice were primed i.p. with $HA_{NY95}$ in alhydrogel and boosted homologously or heterologously with $HA_{N99}$ or $HA_{CA09}$ in alhydrogel as indicated. (**c**) Full time course of anti-HA tag-specific ELISA reactivity (optical density at 1:100 dilution) of the same mice shown in Fig. 3f. Anti-$HA_{NY95}$ (top), $HA_{NC99}$ (middle) and $HA_{CA09}$ (bottom) ELISAs are shown, with Strep (left) and FLAG (right) detection. (**d**) Relationship between primary addiction and antigenic distance for infection and immunization experiments. Graph compiles primary addiction indices from Fig. 3d and 3g. Bars are ordered by amino acid identity between the priming and boosting HAs. P-values are for one-way ANOVA, with % identity as a categorical variable, or Pearson correlation, with (100 – % identity) as a linear variable.

## Extended Data Fig. 5 Neutralization of WH1 and BA.1 pseudoviruses by serum antibody fractions from BA.1 boosted mice.

(**a**) Efficiency of serum fractionation into FLAG and Strep-depleted fractions, as measured by anti-RBD ELISA of input vs. post-depletion samples. Serum was obtained from heterologously immunized mice two weeks after the 3rd immunization, Fig. 4d (8 mice from 2 independent experiments). (**b**) Neutralization of WH1 (left) and BA.1 (right) SARS-CoV-2 S-expressing pseudotyped HIV-1 virus by serum fractions shown in (a). Mean values of technical duplicates are shown. (**c**) WH1 S pseudovirus

NT50 titres for samples in (b). Post-depletion NT50s were normalized to input serum based on the BA.1 RBD ELISA titres (a), by applying a correction equalizing the anti-RBD Strep-titre of the FLAG-depleted fraction to the input, and the anti-RBD FLAG-titre of the Strep-depleted fraction to the input. P-values are for one-tailed paired T test. FC, fold-change. (**d**) Estimating the contribution of *de novo* vs. memory-derived antibodies to excess BA.1 neutralization by the WBB regimen. The contribution of secondary affinity maturation or preferential selection of crossreactive memory B cells by BA.1 boosting, $\Delta S$, is calculated as the difference between $Strep^+$ WBB and total WWW BA.1 neutralization titres (the latter are assumed to be all $FLAG^+$). The contribution of BA.1-specific antibodies induced *de novo* by BA.1 boosting, $F$, is given as the $FLAG^+$ WBB titre. Percent contribution of $F$ to the improved BA.1 neutralization in WBB is calculated as $(F/(F + \Delta S))*100$. Bars and dotted lines represent the medians for each condition, which were used to calculate F and ΔS. The upper dotted line (4.) represents the median neutralization of total WBB antibodies and is shown for reference purposes only. Data in groups 1 and 4 are reproduced from Fig. 4b, data in group 2 and 3 from Fig. 4h.

## Extended Data Fig. 6 Yeast-displayed deep mutational scanning to map mutations that reduce binding of immunized mouse serum.

(**a**) Top: Representative plots of nested FACS gating strategy used for all experiments to select for single yeast cells. Bottom: Gating strategy to select for RBD-expressing single cells (FITC-A vs. FSC-A). (**b**) FACS gating strategy for one of two independent libraries to select cells expressing BA.1 or WH1 RBD mutants with reduced Strep or FLAG antibody binding (cells in blue), as measured by secondary staining with APC-conjugated streptavidin or APC-conjugated anti-FLAG antibody, respectively. Gates were set manually for each sample to capture cells that have a reduced amount of tagged antibody binding for their degree of RBD expression. FACS scatter plots were qualitatively similar between the two libraries. The mouse identifier (#1-4), DMS target library (WH1 or BA.1), and antibody tag (Strep or FLAG) are indicated above each plot. (**c**) Mutation (left)- and site (right)-level correlations of escape scores between two independent biological replicate libraries.

**Extended Data Fig. 7 RBD serum-escape logo plots.**

Deep mutational scanning results of serum collected from 4 heterologously immunized mice, 2 weeks after the 3rd dose (Fig. 4d). Each mutation's "escape fraction" was measured, which ranges from 0 (no cells effect on antibody binding) to 1 (all cells with the mutation have decreased antibody binding). Mouse 4 is not shown in the main text as there were no interpretable peaks in antibody binding to BA.1 libraries for either antibody fraction. Logo plots show the antibody-escape fractions for individual amino-acid mutations at key sites of strong escape. Sites in which BA.1 differs from the WH1 sequence are shown in purple font. All escape scores are shown in Supplementary Table 1 and are available online at https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_OAS/blob/main/results/supp_data/all_raw_data.csv.

## Supplementary information

**Supplementary Table 1**

**Reporting Summary**

## Rights and permissions

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Reprints and Permissions

## About this article

### Cite this article

Schiepers, A., van ’t Wout, M.F.L., Greaney, A.J. *et al.* Molecular fate-mapping of serum antibody responses to repeat immunization. *Nature* **615**, 482–489 (2023). https://doi.org/10.1038/s41586-023-05715-3

- Received: 29 August 2022
- Accepted: 06 January 2023
- Published: 16 January 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05715-3

### Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

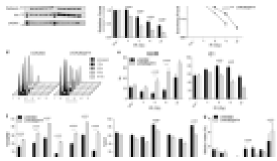
## This article is cited by

- **How your first brush with COVID warps your immunity**
    - Rachel Brazil

  *Nature* (2023)

---

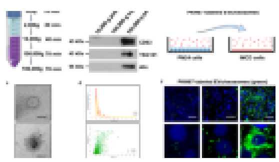
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05715-3



## Fumarate hydratase loss promotes mitotic entry in the presence of DNA damage after ionising radiation

06 September 2018

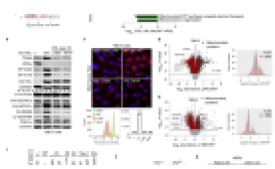
Timothy I. Johnson, Ana S. H. Costa, … Christian Frezza



## Extracellular vesicles and exosomes generated from cystic renal epithelial cells promote cyst growth in autosomal dominant polycystic kidney disease

27 July 2021

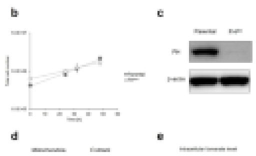
Hao Ding, Linda Xiaoyan Li, … Xiaogang Li



## Impaired oxygen-sensitive regulation of mitochondrial biogenesis within the von Hippel–Lindau syndrome

27 June 2022

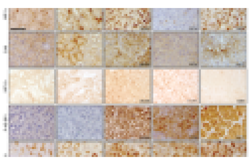
Shuijie Li, Wenyu Li, … Susanne Schlisio



## [Direct and quantitative analysis of altered metabolic flux distributions and cellular ATP production pathway in fumarate hydratase-diminished cells]

03 August 2020

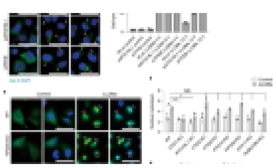
Shingo Noguchi, Hirokazu Ishikawa, … Hiroshi Shimizu



## [HIF-1α and HIF-2α differently regulate tumour development and inflammation of clear cell renal cell carcinoma in mice]

17 August 2020

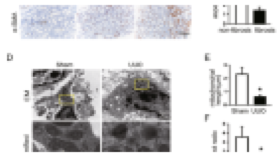
Rouven Hoefflin, Sabine Harlander, … Ian J. Frew



## [LC3 lipidation is essential for TFEB activation during the lysosomal damage response to kidney injury]

28 September 2020

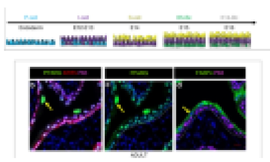
Shuhei Nakamura, Saki Shigeyama, … Tamotsu Yoshimori



## [Drp1-mediated mitochondrial fission promotes renal fibroblast activation and fibrogenesis]

16 January 2020

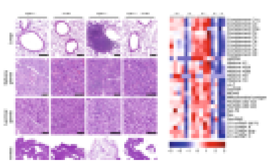
Yating Wang, Miaoqing Lu, … Haiping Mao



## [Pparg promotes differentiation and regulates mitochondrial gene expression in bladder epithelial cells]

09 October 2019

Chang Liu, Tiffany Tate, … Cathy Lee Mendelsohn



## [IRGM1 links mitochondrial quality control to autoimmunity]

28 January 2021

Prashant Rai, Kyathanahalli S. Janardhan, … Michael B. Fessler

- Article
- Open Access
- Published: 08 March 2023

# Fumarate induces vesicular release of mtDNA to drive innate immunity


- Vincent Zecchini,
- Vincent Paupe,
- Irene Herranz-Montoya,
- Joëlle Janssen,
- Inge M. N. Wortel,
- Jordan L. Morris,
- Ashley Ferguson,
- Suvagata Roy Chowdury,
- Marc Segarra-Mondejar,
- Ana S. H. Costa,
- Gonçalo C. Pereira,
- Laura Tronci,
- Timothy Young,
- Efterpi Nikitopoulou,
- Ming Yang,
- Dóra Bihary,
- Federico Caicci,
- Shun Nagashima,
- Alyson Speed,
- Kalliopi Bokea,
- Zara Baig,
- Shamith Samarajiwa,
- Maxine Tran,
- Thomas Mitchell,
- Mark Johnson,
- Julien Prudent &
- …
- Christian Frezza

Show authors



*Nature* **volume 615**, pages 499–506 (2023)

- 20k Accesses
- 272 Altmetric
- Metrics details

## Abstract

Mutations in fumarate hydratase (FH) cause hereditary leiomyomatosis and renal cell carcinoma[1]. Loss of FH in the kidney elicits several oncogenic signalling cascades through the accumulation of the oncometabolite fumarate[2]. However, although the long-term consequences of FH loss have been described, the acute response has not so far been investigated. Here we generated an inducible mouse model to study the chronology of FH loss in the kidney. We show that loss of FH leads to early alterations of mitochondrial morphology and the release of mitochondrial DNA (mtDNA) into the cytosol, where it triggers the activation of the cyclic GMP–AMP synthase (cGAS)–stimulator of interferon genes (STING)–TANK-binding kinase 1 (TBK1) pathway and stimulates an inflammatory response that is also partially dependent on retinoic-acid-inducible gene I (RIG-I). Mechanistically, we show that this phenotype is mediated by fumarate and occurs selectively through mitochondrial-derived vesicles in a manner that depends on sorting nexin 9 (SNX9). These results reveal that increased levels of intracellular fumarate induce a remodelling of the mitochondrial network and the generation of mitochondrial-derived vesicles, which allows the release of mtDNAin the cytosol and subsequent activation of the innate immune response.

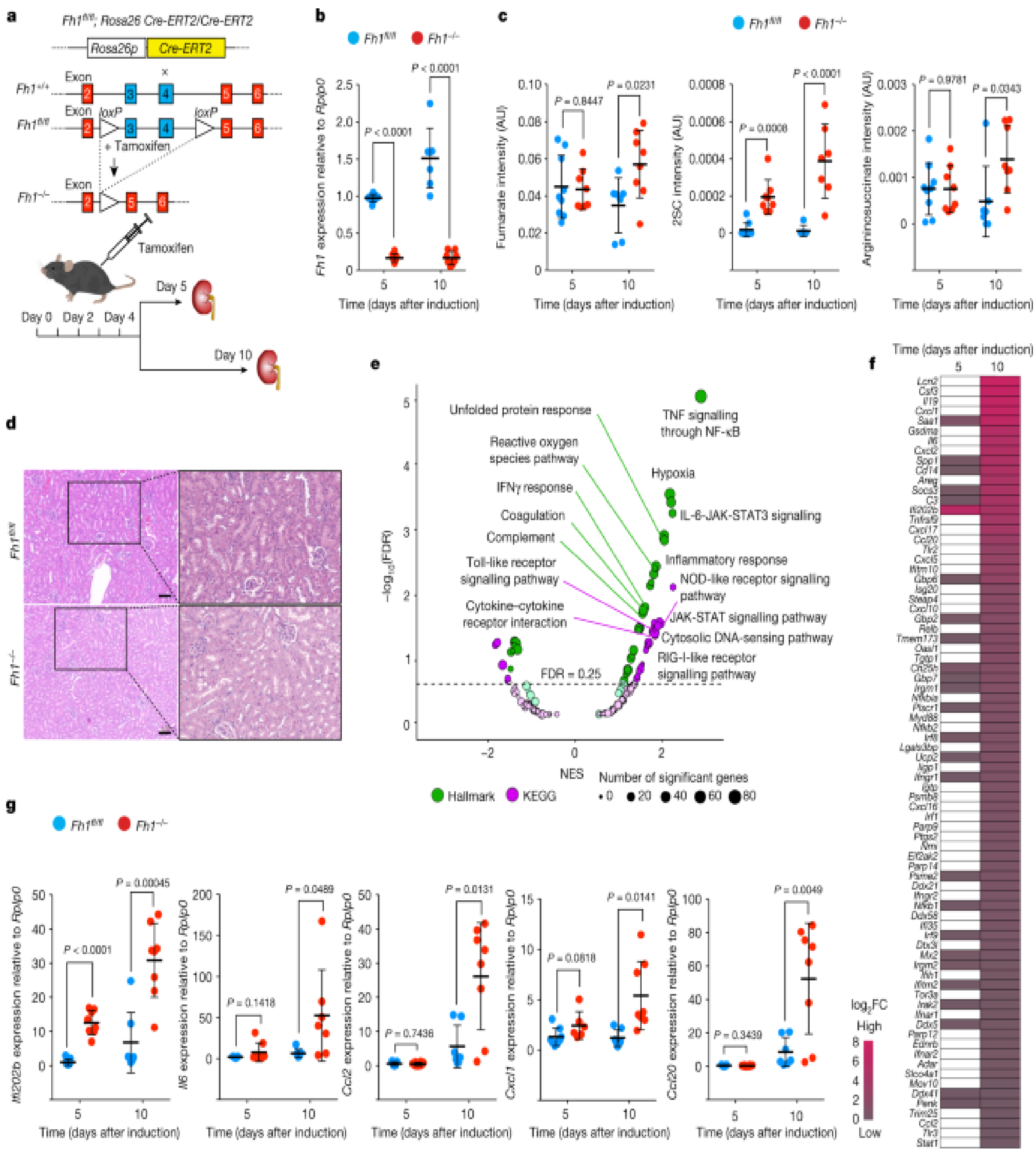
## Main

Fumarate hydratase (FH), a metabolic enzyme of the tricarboxylic acid (TCA) cycle that catalyses the reversible conversion of fumarate to malate, has been identified as a bona fide tumour suppressor[2]. Loss of FH predisposes to hereditary leiomyomatosis and renal cell carcinoma (HLRCC), a syndrome that is characterized by an aggressive form of kidney cancer. A hallmark of both FH-deficient cells and tumours is the aberrant

accumulation of fumarate[3], which has been shown to drive malignant transformation and tumour progression through the activation of a series of oncogenic cascades[2]. However, when analysed in tumours or other cellular models, all of these signalling cascades are already simultaneously activated and are likely to be intertwined, which makes it difficult to elucidate their hierarchy, cross-talk and causal contribution to the disease.

Here we generated an inducible transgenic mouse model to investigate the chronology of the loss of *Fh1*, the mouse orthologue of human *FH*, in the adult kidney. By crossing a previously generated $Fh1^{fl/fl}$ strain[4] with mice that contain a ubiquitously expressed tamoxifen-inducible Cre recombinase at the *Rosa26* locus[5], we obtained a tamoxifen-inducible $Fh1^{fl/fl}$ strain (Fig. 1a). Treating $Fh1^{fl/fl}$ mice with tamoxifen resulted in the efficient excision of *Fh1* exons 3 and 4 (hereafter termed $Fh1^{-/-}$), compared to the control strain ($Fh1^{+/+}$) (Fig. 1b and Extended Data Fig. 1a,b). Loss of FH causes profound metabolic changes, including the aberrant intracellular accumulation of argininosuccinate[6] and of products of the reaction by which fumarate is added to the thiol groups of free cysteine and to the thiol groups of proteins[7], generating *S*-(2-succinyl)cysteine (2SC) and succinated proteins, respectively. Both reactions buffer the excess of fumarate production, which eventually occurs when these and other buffering systems have reached capacity (Extended Data Fig. 1c). We observed an early increase of 2SC at day 5, followed by an increase of all these metabolic markers, including fumarate, at day 10 in $Fh1^{-/-}$ kidneys (Fig. 1c). Macroscopically, there were no gross morphological differences between wild-type and *Fh1*-deficient kidneys (Fig. 1d).

**Fig. 1: Loss of *Fh1* in the adult mouse kidney triggers an early inflammatory response.**



**a**, Genome-editing strategy for generating inducible *Fh1* knockout alleles. *Rosa26-Cre-ERT2* mice carry the tamoxifen-responsive *Cre* recombinase transgene downstream of the *Rosa26* promoter[5]. *Fh1*$^{fl/fl}$ mice[13] contain two *loxP* sites flanking *Fh1* exons 3 and 4. Intraperitoneal injection of tamoxifen in the adult mouse induces the nuclear translocation of the ubiquitously expressed Cre–ERT2 fusion protein, resulting in the excision of the genomic fragment located between the *loxP* sites to generate *Fh1* null alleles (*Fh1*$^{-/-}$). Mice of about 90 days of age are treated with tamoxifen and the kidneys

are collected for analysis at day 5 and day 10 after injection. **b**, *Fh1* mRNA expression levels in wild-type control ($Fh1^{+/+}$) and *Fh1*-deficient ($Fh1^{-/-}$) adult mouse kidneys, measured by qRT–PCR. **c**, Metabolite abundance (normalized peak ion intensity, arbitrary units (AU)) in $Fh1^{+/+}$ and $Fh1^{-/-}$ adult mouse kidney measured by liquid chromatography–mass spectrometry (LC–MS). **d**, Haematoxylin and eosin (H&E) staining of $Fh1^{+/+}$ and $Fh1^{-/-}$ adult mouse kidney at day 10 after induction. Scale bars, 100 µm. **e**, Volcano plots of the GSEA, highlighting the differentially regulated pathways in $Fh1^{-/-}$ versus $Fh1^{+/+}$ kidney tissue at day 10 after induction. FDR, false discovery rate; NES, normalized enrichment score. **f**, Heat map showing upregulated inflammation-related genes in $Fh1^{-/-}$ versus $Fh1^{+/+}$ kidney tissue at day 5 and day 10 after induction (white cells mean no difference between the two groups). **g**, Expression levels of ISGs in $Fh1^{-/-}$ versus $Fh1^{+/+}$ kidney tissue at day 5 and day 10 after induction, measured by qRT–PCR. Data are mean ± s.e.m. **b**,**c**,**g**, $n$ = minimum 8 mice in each group, Student's $t$-test corrected for multiple comparisons with the Holm–Sidak method.

Source data

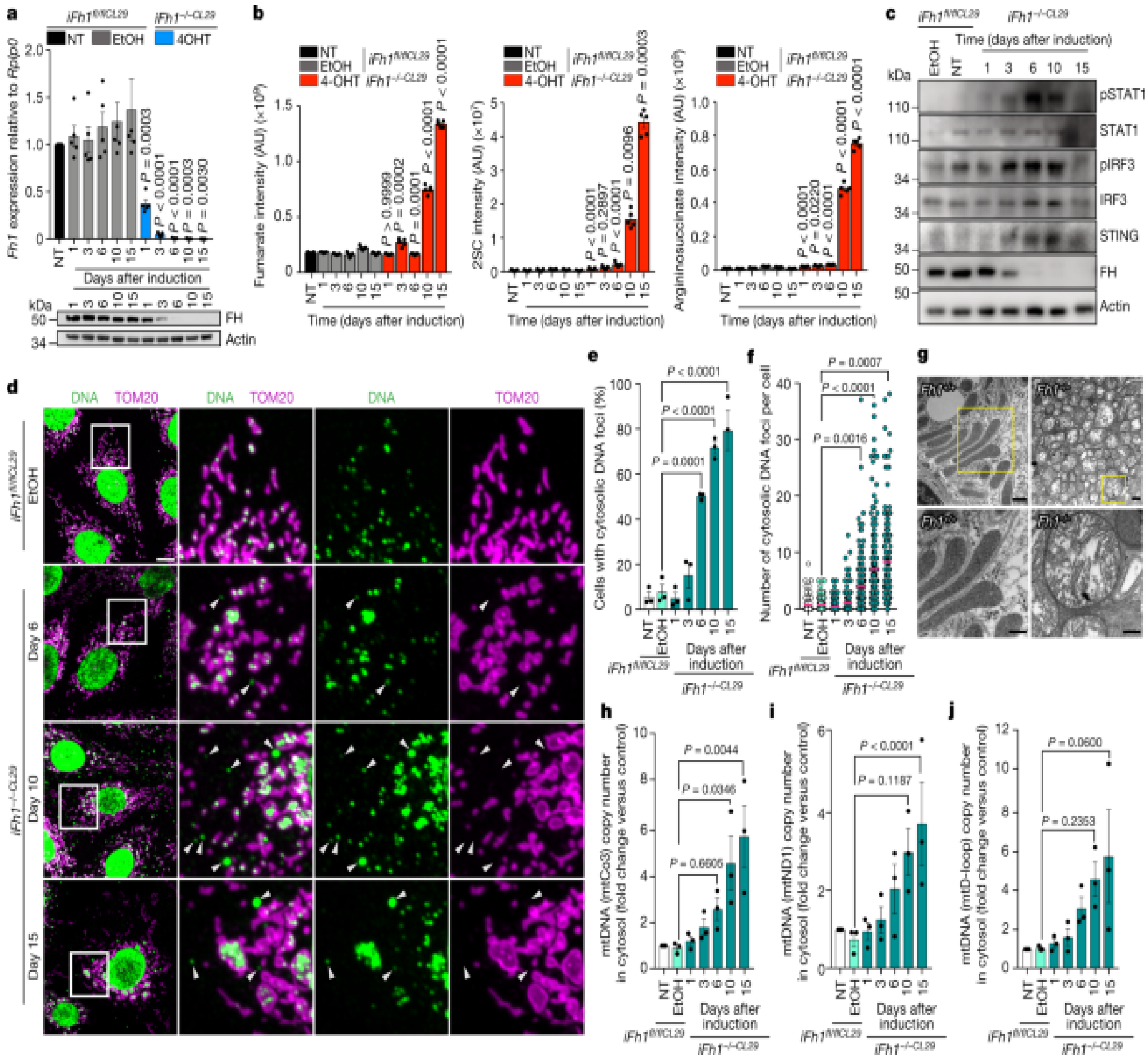
## Loss of *Fh1* triggers an inflammatory response

To investigate the acute response to the loss of *Fh1*, we performed a transcriptomic analysis of the mouse kidney (Fig. 1a and Supplementary Table 1). The most prevalent altered biological processes were linked to pathways that are involved in inflammation and the innate immune response (Fig. 1e and Supplementary Tables 2 and 3). We also observed an increase in interferon-stimulated genes (ISGs), with *Ifi202b* being the most upregulated transcript (Fig. 1f, Extended Data Fig. 1d and Supplementary Table 1), and found that this was associated with an upregulation of pro-inflammatory cytokines and chemokines (Fig. 1f,g). Deconvolution of the transcriptomics data revealed that immune cells contributed to a low fraction of the signal that was observed (Extended Data Fig. 1e). In addition, although we confirmed the presence of succinated proteins, there were no changes in the expression of CD14 (a marker of immune cells) in $Fh1^{-/-}$ kidney tissue

(Extended Data Fig. 1f), ruling out infiltrating immune cells as the cause of the immune signature observed.

To confirm the cell-autonomous nature of the phenotype, we generated epithelial cell lines from *Fh1*$^{fl/fl}$ mouse kidneys[8], hereafter referred to as inducible *Fh1* clones 29 or 33 (*iFh1*$^{-/-CL29}$ or *iFh1*$^{-/-CL33}$ after recombination and *Fh1* loss) (Extended Data Fig. 2a). Treating *iFh1*$^{fl/flCL29}$ and *iFh1*$^{fl/flCL33}$ cells with 4-hydroxytamoxifen (4-OHT) resulted in a time-dependent decrease in the mRNA and protein levels of *Fh1*, accompanied by reductions in the mitochondrial membrane potential and respiration; it also induced an early increase in the metabolic markers 2SC and argininosuccinate, followed by the accumulation of fumarate, and transcriptional hallmarks of *Fh1* loss (Fig. 2a,b, Extended Data Fig. 2b–d and Supplementary Fig. 1a–d).

**Fig. 2: Activation of the cGAS–STING pathway in *Fh1*-deficient cells is triggered by cytosolic mtDNA.**




**a**, qRT–PCR (top) and immunoblots (bottom) showing the expression levels of *Fh1* (top) or FH protein (bottom) in inducible *iFh1* epithelial kidney cell lines clones 29 (*iFh1*$^{fl/flCL29}$) that were not treated (NT) or were treated with either vehicle (ethanol; EtOH) or 4-OHT (*iFh1*$^{-/-CL29}$) for the indicated period of time. $n = 5$ independent experiments. **b**, Relative abundance (normalized peak ion intensity) of fumarate (left), 2SC (middle) and argininosuccinate (right) in *iFh1*$^{CL29}$ cells measured by LC–MS. $n = 5$ independent experiments. **c**, Immunoblots of specified proteins in *iFh1*$^{CL29}$ cells. **d**, Representative confocal images of mitochondrial morphology (TOM20) and DNA foci (DNA) in *iFh1*$^{CL29}$ cells. White arrowheads indicate cytosolic DNA foci. Scale bar, 10 µm. **e**,**f**, Percentage of *iFh1*$^{CL29}$ cells showing cytosolic DNA foci (**e**) and number of cytosolic DNA foci per

cell (**f**) from **d**. $n = 3$ independent experiments. **g**, TEM images of mitochondria from *Fh1*$^{+/+}$ and *Fh1*$^{-/-}$ adult mouse kidney tissue. Scale bars, 1 µm (top); 500 nm (bottom). **h**–**j**, Quantification of mtDNA copy number by ddPCR using a mtCo3 (**h**), mtND1 (**i**) or mtD-loop (**j**) probe, from isolated cytosolic fractions of *iFh1*$^{CL29}$ cells at day 1 to day 15 after induction. $n = 3$ independent experiments. Data are mean ± s.e.m. **a**,**b**, Student's *t*-test corrected for multiple comparisons with the Holm–Sidak method; *P* values are indicated above each condition and relative to the corresponding vehicle-treated control; **e**,**f**,**h**–**j**, one-way ANOVA with Tukey's multiple comparisons test.

Source data

We then investigated the pathways involved in the activation of inflammation. The binding of conserved features of invading pathogens or that of host components released as a result of cell or tissue damage converges on a central kinase, TBK1, which in turn activates interferon-regulatory factors (IRFs) and leads to the production of pro-inflammatory cytokines, chemokines and type I and III interferons (IFNs)[9] (Extended Data Fig. 2e). Consistent with the activation of TBK1, we observed a time-dependent phosphorylation of IRF3 and the downstream effector STAT1 in *iFh1*$^{-/-CL29}$ cells, accompanied by an upregulation of STING (Fig. 2c) and an increase in the transcription of ISGs in both *iFh1*$^{-/-}$ clones (Extended Data Fig. 2f and Supplementary Fig. 1e).

Altogether, these results show that the loss of *Fh1* triggers a cell-autonomous innate immune response that results in the upregulation of ISGs both in vivo and in vitro.

# Loss of *Fh1* induces the release of cytosolic mtDNA

Previous studies have shown that mtDNA and mitochondrial RNA (mtRNA) have a role as potent activators of nucleic-acid-sensing pathways, leading to the activation of different inflammatory pathways[10]. We observed a time-dependent increase in the number of cells containing cytosolic DNA foci,

which matched the dynamics of the loss of *Fh1* expression (Fig. 2d–f and Supplementary Fig. 1f,g) and coincided with a progressive remodelling of mitochondrial morphology that was characterized by swollen and swollen-elongated mitochondria (Fig. 2d, Extended Data Fig. 2g–i and Supplementary Fig. 1h–j). A transmission electron microscopy (TEM) analysis of *Fh1*-deficient kidneys (Fig. 2g) and other mouse (Extended Data Fig. 2j) and human[11] FH-deficient lines identified similar anomalies in mitochondrial ultrastructure. Therefore, we hypothesized that the source of the DNA foci could be mtDNA leaking into the cytosol. To test this, we quantified the mtDNA copy number in isolated cytosolic fractions using droplet digital PCR (ddPCR). We found a time-dependent increase in mtDNA copy number in the cytosol of both $iFh1^{-/-CL29}$ and $iFh1^{-/-CL33}$ cell lines (Fig. 2h–j and Supplementary Fig. 1k–m).

To determine whether this phenotype was maintained over time, we resorted to the chronic model of *Fh1* loss that we previously generated[12] (hereafter termed $cFh1^{fl/fl}$ for the control line and $cFh1^{-/-CL1}$ and $cFh1^{-/-CL19}$ for the two Fh1-deficient clones) (Supplementary Fig. 2a). Similar to $iFh1^{-/-}$ cells, both $cFh1^{-/-}$ cell lines exhibited anomalies in mitochondrial morphology (Extended Data Fig. 3a–e and Supplementary Fig. 2b). In addition, both $cFh1^{-/-}$ cell lines showed the release of mtDNA into the cytosol, activation of the STING–TBK1–IRF3 cascade and upregulation of the expression of ISGs (Extended Data Fig. 3f–l). All aspects of this phenotype were fully rescued by the re-expression of exogenous *Fh1* (Extended Data Fig. 3c–l). Together, these results indicate that the loss of *Fh1* induces a chronic release of mtDNA into the cytosol, which triggers a persistent inflammatory response.

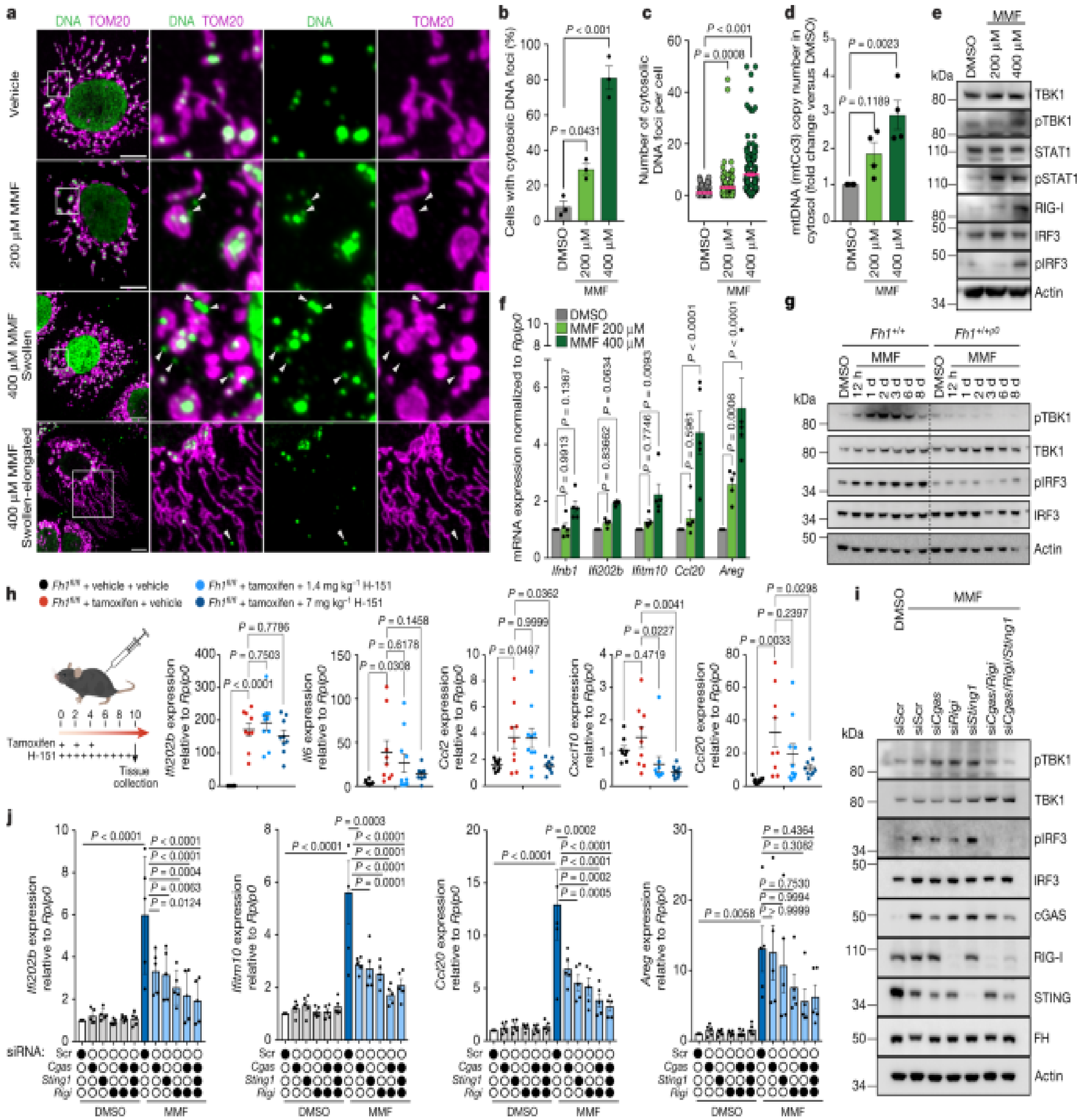
We also showed that cells deficient for the B subunit of succinate dehydrogenase (SDH)[13], another TCA-cycle enzyme, exhibited no significant enrichment in immune-related processes, no major changes in mitochondrial morphology, low levels of cytosolic mtDNA and low transcription levels of *Ifnb1*, ISGs and *Hmox1* (Supplementary Fig. 3a–k). Furthermore, expressing NDI1 (ref. [14]) to restore the defects in the electron transport chain that were observed in *Fh1*-deficient cells[15] did not rescue any phenotype (Supplementary Fig. 4).

Together, these results suggest that the loss of FH activity, rather than an overall impairment of the TCA cycle or mitochondrial metabolism, is responsible for the activation of the inflammatory response.

## Fumarate phenocopies *Fh1* loss

We then tested whether fumarate could be the main driver behind this effect (Extended Data Fig. 1c), by treating *cFh1*$^{fl/fl}$ cells with a cell-permeable derivative of fumarate, monomethyl fumarate (MMF), which is known to increase the cellular levels of fumarate without substantially affecting its reactivity[16] (Supplementary Fig. 5a). Treatment with MMF phenocopied the effects of *Fh1* loss, inducing the dose- and time-dependent expression of hallmarks of *Fh1* loss, 2SC and fumarate at levels similar to those in *iFh1*$^{-/-}$ cells, early mitochondrial accumulation of succinated proteins, mitochondrial network remodelling, cytosolic mtDNA release, activation of the TBK1 cascade and expression of ISGs (Fig. 3a–f and Extended Data Figs. 4 and 5). Expression of cytosolic *Fh1*, which partially restores the oxygen consumption rate (OCR) and the levels of fumarate, but still leads to increased mitochondrial succination, only partially rescued the phenotype (Extended Data Fig. 6), suggesting that increased mitochondrial fumarate levels are primarily responsible for mtDNA release. Treatment with a cell-permeable form of succinate, which accumulates in SDH-deficient cells and tissues and to a much lesser extent also in FH-deficient conditions[17], had no such effect (Extended Data Fig. 7a–d and Supplementary Fig. 5b–f). Similarly, treatment with the TCA-cycle-derived metabolites α-ketoglutarate or 2-hydroxyglutarate had no effect compared to MMF treatment (Extended Data Fig. 7e–g and Supplementary Fig. 5g).

**Fig. 3: Fumarate induces a remodelling of mitochondrial morphology and the release of mtDNA.**



**a**–**f**, Chronic *Fh1* kidney cells (*cFh1*$^{fl/fl}$) were treated with 200 μM or 400 μM MMF or vehicle (dimethyl sulfoxide; DMSO) for 8 days. $n = 3$ independent experiments (unless otherwise specified). In **g**, *iFh1*$^{fl/flCL29}$ cells were treated with 400 μM MMF or DMSO for 8 days. **a**, Representative confocal images of mitochondrial morphology (TOM20) and DNA foci (DNA) in *cFh1*$^{fl/fl}$ cells. White arrowheads indicate cytosolic DNA foci. Scale bars, 10 μm. **b**,**c**, Percentage of *cFh1*$^{fl/fl}$ cells showing cytosolic DNA foci (**b**) and number of cytosolic DNA foci per cell (**c**) from **a**. **d**, Quantification of mtDNA copy number by ddPCR using a mtCo3 probe,

from isolated cytosolic fractions of *cFh1*$^{fl/fl}$ cells. **e**, Immunoblots of specified proteins in *cFh1*$^{fl/fl}$ cells. **f**, ISG expression in *cFh1*$^{fl/fl}$ cells measured by qRT–PCR. $n = 5$ independent experiments. **g**, Immunoblots of specified proteins in *iFh1*$^{fl/flCL29}$ and mtDNA-depleted *iFh1*$^{fl/flCL29\rho0}$ cells. **h**, mRNA expression of a panel of ISGs in mouse kidney tissue treated with the STING inhibitor H-151, measured by qRT–PCR. $n = 9$ mice per group. **i**,**j**, Immunoblots of specified proteins (**i**) and ISG expression measured by qRT–PCR (**j**) in *cFh1*$^{fl/fl}$ cells treated with 400 µM MMF or DMSO for 8 d, and transfected with the indicated siRNAs (Scr, scramble). $n = 5$ independent experiments. Data are mean ± s.e.m. **b**–**d**,**f**,**j**, One-way ANOVA with Tukey's multiple comparison test; **h**, one-way ANOVA with Dunnett's multiple comparison test.

Source data

To further confirm the role of mtDNA release in our model, we generated mtDNA-depleted *iFh1*$^{fl/flCL29\rho0}$ (that is, *Fh1*$^{+/+\rho0}$) cells (Extended Data Fig. 8a). MMF-treated *Fh1*$^{+/+\rho0}$ cells did not show any increase in phosphorylated TBK1 (pTBK1) and pIRF3 (Fig. 3g), or other downstream ISG targets (Extended Data Fig. 8b).

These results show that fumarate and cytosolic mtDNA are the trigger and the ensuing stimulus, respectively, that account for the activation of the innate immune response that is observed in *Fh1*-deficient cells.

# Fumarate induces activation of cGAS and RIG-I

The presence of cytosolic mtDNA and the activation of the TBK1–IRF3 cascade in *Fh1*-deficient cells suggests that the cGAS–STING pathway might be involved in eliciting the observed inflammatory phenotype. We observed increased cytosolic cGAS in MMF-treated cells (Extended Data Fig. 8c,d) and showed that the pharmacological inhibition of cGAS in MMF-treated cells, *iFh1*$^{-/-CL29}$ and *cFh1*$^{-/-CL1}$ cells using RU.521 (ref. [18]) lowered the expression levels of ISGs (Extended Data Fig. 8e–g). The in vivo pharmacological inhibition of STING also partially rescued the upregulation of ISGs (Fig. 3h and Extended Data Fig. 8h), indicating that the

cGAS–STING pathway is a crucial part of the inflammatory response induced both in vitro and in vivo.

Notably, we found that silencing the cytosolic RNA sensor RIG-I (*Rigi*), but not its adaptor protein *Mavs* or the RNA sensor *Mda5* (also known as *Ifih1*), also reduced the levels of TBK1 and IRF3 phosphorylation in $iFh1^{-/-CL29}$ cells (Extended Data Fig. 8i and Supplementary Fig. 6a). Similarly, although the individual silencing of *Sting1*, *Cgas* or *Rigi* in MMF-treated cells had a mild effect, their simultaneous knockdown enhanced the rescue of TBK1 and IRF3 phosphorylation and ISG expression (Fig. 3i,j and Supplementary Fig. 6b).

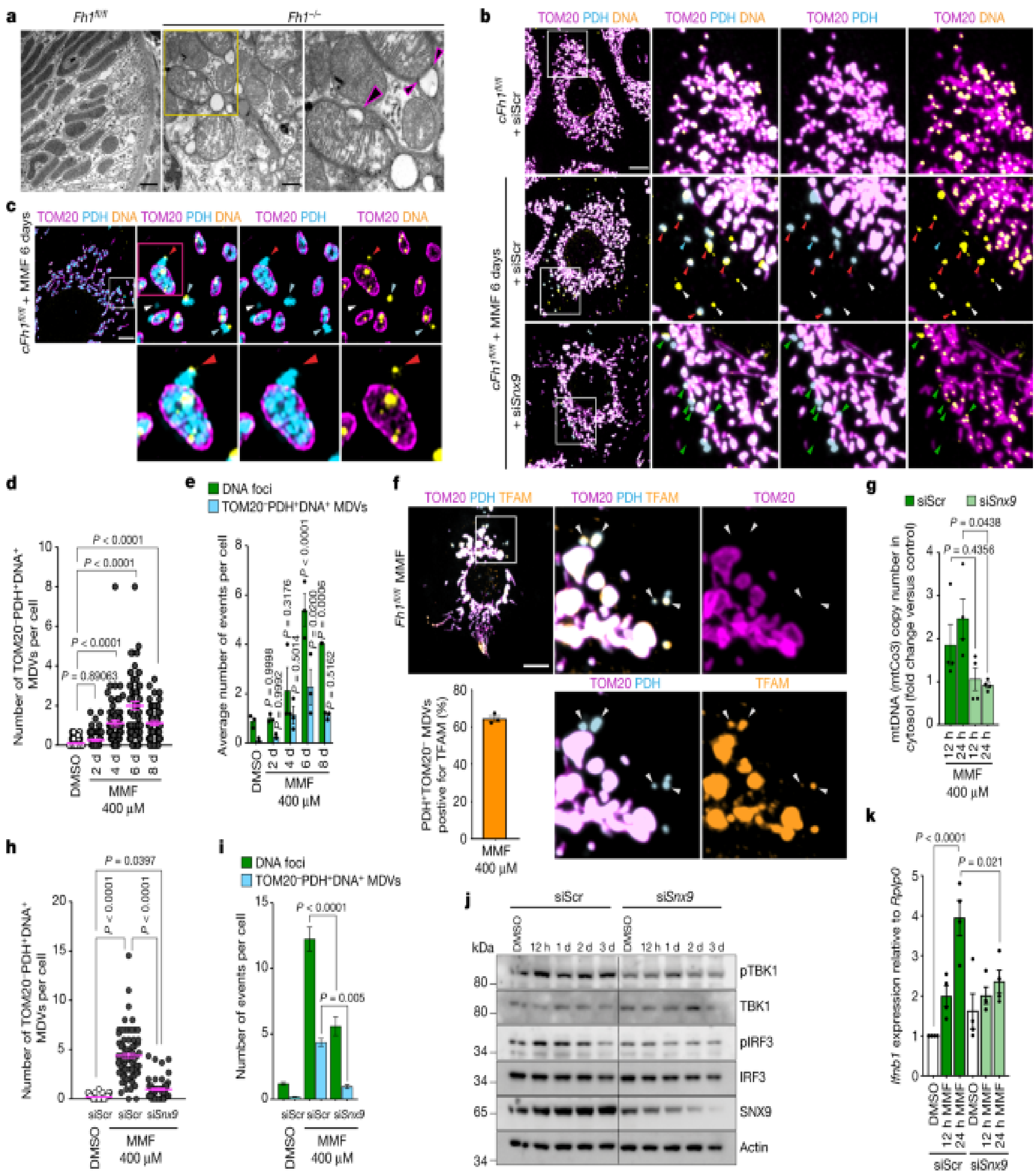
Together, these results indicate that a complex network that includes at least two cytoplasmic nucleic-acid sensors—cGAS and RIG-I—activates the TBK1–IRF3-dependent response after the release of mtDNA into the cytosol.

## mtDNA is released through SNX9-dependent MDVs

To gain insight into the mechanism of mtDNA release mediated by FH deficiency, we used $cFh1^{-/-CL1}$ cells to perform a targeted small interfering RNA (siRNA) screen. The double silencing of *Bax* and *Bak1*, which are involved in mtDNA release[19], did not rescue cytosolic mtDNA release (Extended Data Fig. 9a–d and Supplementary Fig. 7). Of note, deletion of *Fh1* did not lead to the release of cytochrome *c* or the recruitment of BAX to mitochondrial membranes (Extended Data Fig. 9e,f). Knockdown of *Vdac1*, which has been shown to facilitate mtDNA release[20], was also unable to rescue the phenotype (Extended Data Fig. 9a–d and Supplementary Fig. 7). Knockdown of the main regulators of mitochondrial fusion and fission[21], or *Cgas*, *Sting1* and *Rigi*, did not affect mtDNA release (Extended Data Fig. 9a–d and Supplementary Fig. 7). Finally, the levels of TFAM (transcription factor A, mitochondrial) were unchanged in our models (Supplementary Fig. 8), ruling out a role for TFAM depletion in fumarate-dependent mtDNA release[22].

As an alternative way to communicate with other organelles and transport mitochondrial content without affecting the integrity of membranes, mitochondria can generate small-vesicle carriers known as mitochondrial-derived vesicles (MDVs)[23]. TEM imaging of $Fh1^{-/-}$ mouse kidneys revealed double-membrane-bound, low-density-content vesicles protruding from mitochondria, similar to budding MDVs[24] (Fig. 4a). The role of SNX9 (an endocytic accessory protein) and RAB9 in the release of MDVs was previously reported[25]. Here, the silencing of *Snx9*, but not *Rab9*, significantly reduced the release of mtDNA in both $cFh1^{-/-CL1}$ and $cFh1^{-/-CL19}$ cells (Extended Data Figs. 9a–d and 10a–c and Supplementary Fig. 7). Loss of *Snx9* also decreased the activation of TBK1–IRF3 phosphorylation, without restoring defects in mitochondrial morphology (Extended Data Fig. 10d–f). Finally, these results were confirmed in $iFh1^{-/-}$ cells at 10 and 15 days after induction (Extended Data Fig. 10g–l and Supplementary Fig. 9a). Together, these results indicate that SNX9 has a crucial role in the inflammatory response that is induced by cytosolic mtDNA release.

**Fig. 4: mtDNA is conveyed to the cytosol by MDVs.**



For all panels: $n = 3$ independent experiments (unless otherwise specified). In **b**,**c**,**f**,**h**,**i**, *cFh1*$^{fl/fl}$ cells were treated with 400 μM MMF or vehicle (DMSO) for 6 days. For immunofluorescence experiments, mitochondria or MDVs were labelled using anti-TOM20, anti-PDH and anti-TFAM antibodies and DNA with an anti-DNA antibody. **a**, TEM images of *Fh1*$^{+/+}$ and *Fh1*$^{-/-}$ mouse kidney showing potential vesicle budding at the mitochondrial surface (black arrowheads). Scale bars, 1 μm. **b**,

Representative confocal images of *cFh1*$^{fl/fl}$ cells transfected with the indicated siRNAs. Red arrowheads: $TOM20^{-}PDH^{+}DNA^{+}$ MDVs; blue arrowheads: $TOM20^{-}PDH^{+}DNA^{-}$ MDVs; white arrowheads: cytosolic DNA foci; green arrowheads: stalled $TOM20^{-}PDH^{+}DNA^{+}$ MDV budding events. Scale bar, 10 µm. **c**, Representative N-structured illumination microscopy (N-SIM) super-resolution images of MMF-treated *cFh1*$^{fl/fl}$ cells. Red arrowheads: $TOM20^{-}PDH^{+}DNA^{+}$ MDV budding event; blue arrowheads: $TOM20^{-}PDH^{+}DNA^{+}$-released MDVs. Scale bar, 5 µm. **d**,**e**, Quantification of $TOM20^{-}PDH^{+}DNA^{+}$ MDVs (**d**), and compared to cytosolic DNA foci (**e**) in *cFh1*$^{fl/fl}$ cells; *P* values are relative to the corresponding DMSO-treated control. **f**, Representative confocal images of MMF-treated *cFh1*$^{fl/fl}$ cells. White arrowheads: $TOM20^{-}PDH^{+}TFAM^{+}$ MDVs. Scale bar, 10 µm. Bottom left, histogram showing the quantification of $TOM20^{-}PDH^{+}$ MDVs that are positive for TFAM. **g**, Quantification of mtDNA copy number by ddPCR using a mtCo3 probe from isolated cytosolic fractions of MMF-treated *cFh1*$^{fl/fl}$ cells (12 h or 24 h), and pre-transfected with the indicated siRNAs. $n = 4$ independent experiments. **h**,**i**, Quantification of $TOM20^{-}PDH^{+}DNA^{+}$ MDVs (**h**) and compared to cytosolic DNA foci (**i**) in MMF-treated *cFh1*$^{fl/fl}$ cells and pre-transfected with the indicated siRNAs. **j**,**k**, Immunoblots of specified proteins (**j**) and *Ifnb1* mRNA expression measured by qRT–PCR (**k**) in *cFh1*$^{fl/fl}$ cells transfected with the indicated siRNAs and treated with MMF for the indicated period of time. Data are mean ± s.e.m. **d**,**g**–**i**,**k**, One-way ANOVA with Tukey's multiple comparisons test; **e**, two-way ANOVA with Tukey's multiple comparisons test.

Source data

One of the hallmarks of MDVs is their cargo selectivity[23]. We observed an increase in vesicles that were positive for the matrix protein pyruvate dehydrogenase (PDH) but negative for the outer-mitochondrial-membrane protein TOM20 ($TOM20^{-}PDH^{+}$) in MMF-treated cells (Fig. 4b–e and Extended Data Fig. 11a–c). Notably, most of these vesicles also contained DNA ($TOM20^{-}PDH^{+}DNA^{+}$) (Fig. 4d,e and Extended Data Fig. 11a,b). Microscopy analysis (Fig. 4f and Extended Data Fig. 11d) also revealed the presence of the nucleoid-associated protein TFAM inside these

TOM20$^{-}$PDH$^{+}$ and TOM20$^{-}$PDH$^{+}$DNA$^{+}$ MDVs, confirming the mitochondrial origin of the cytosolic DNA foci. Lattice structured illumination microscopy (SIM) analysis showed that these TOM20$^{-}$PDH$^{+}$DNA$^{+}$ MDVs exhibited, on average, a Feret's diameter of 0.45 µm and an area of 0.2 $\mu m^2$ (Extended Data Fig. 11e–g and Supplementary Fig. 9b). We also confirmed the presence of TOM20$^{-}$PDH$^{+}$DNA$^{+}$ MDVs from as early as day 3 after the loss of *Fh1* in $iFh1^{-/-CL29}$ cells (Extended Data Fig. 11h,i), corroborating the early activation of the ISG response that was observed in this inducible model. Finally, treating $Fh1^{+/+\rho0}$ cells with MMF did not reveal the presence of TOM20$^{-}$PDH$^{+}$ MDVs (Extended Data Fig. 12a,b), indicating that the presence of mtDNA is necessary for the formation of MDVs triggered by fumarate.

Silencing *Snx9* before treatment with MMF resulted in a reduction not only in mtDNA release in the cytosol (Fig. 4b,g and Extended Data Fig. 12c), but also in the number of TOM20$^{-}$PDH$^{+}$DNA$^{+}$ MDVs in the cytosol (Fig. 4h,i), and was accompanied by the detection of arrested budding vesicles at the mitochondrial membrane (Fig. 4b). Finally, blocking the release of MDVs also affected the activation of the nucleic-acid-sensing pathway, as silencing *Snx9* mitigated the activation of TBK1–IRF3 and the expression of *Ifnb1* and ISGs that was observed after treatment with MMF (Fig. 4j,k and Extended Data Fig. 12d).

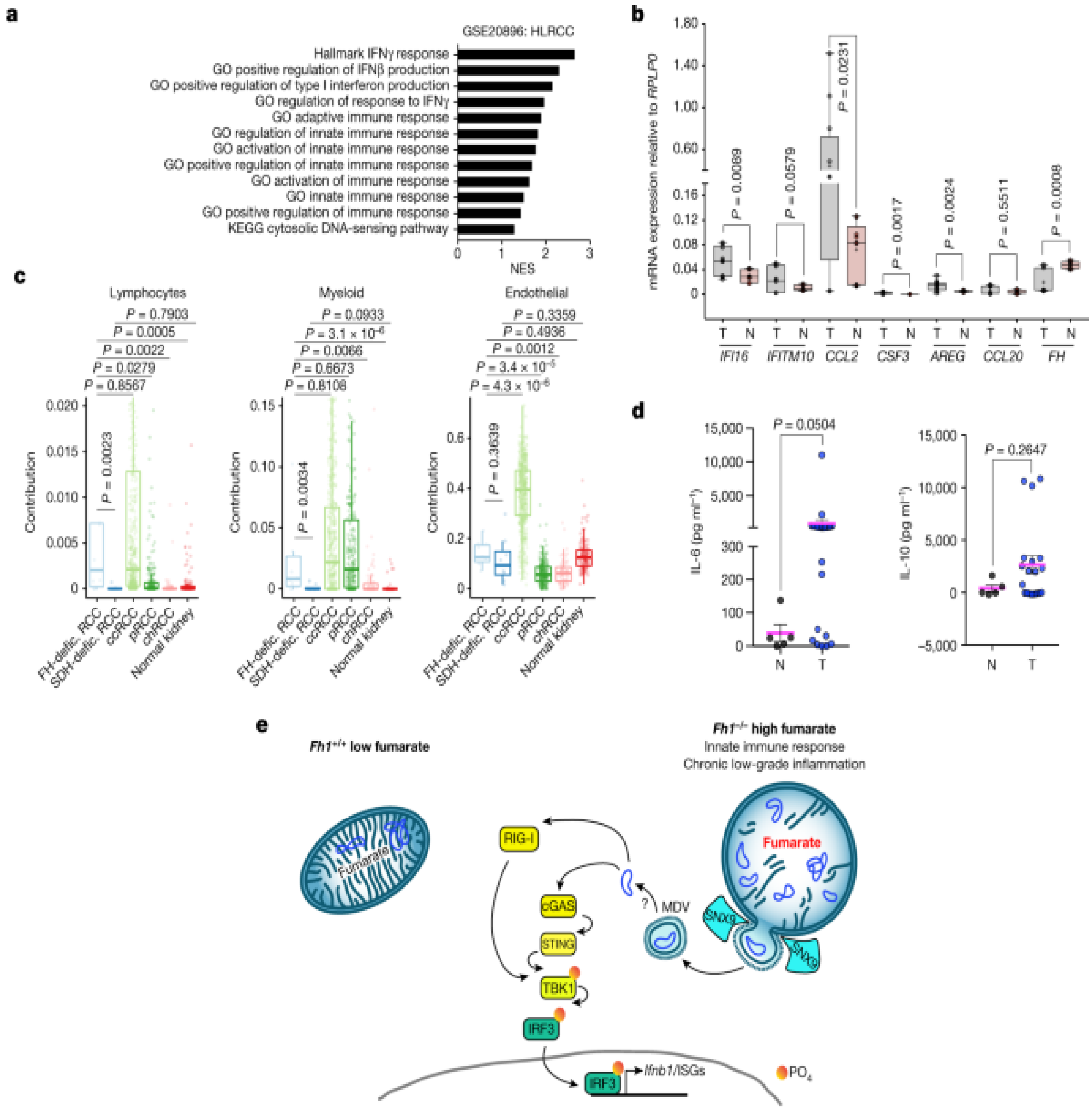
Altogether, these results indicate that the accumulation of fumarate drives the SNX9- and MDV-dependent release of mtDNA into the cytosol to trigger innate immunity.

## Loss of FH causes inflammation in renal cancer

We next investigated whether this inflammatory response could be relevant in patients with HLRCC. Gene set enrichment analysis (GSEA) of a published HLRCC gene expression dataset revealed a chronic inflammatory response, with an activation of innate immunity and DNA-sensor pathways similar to that observed in vivo (Fig. 5a), and quantitative PCR with reverse transcription (qRT–PCR) showed an upregulation of the same key markers

in *FH*-deficient tumours (Fig. 5b). We also estimated the cellular composition of bulk RNA-sequencing datasets of *FH*-deficient renal cell carcinomas (RCCs), *SDH*-deficient RCCs and common RCC subtypes from The Cancer Genome Atlas (TCGA) by deconvolution of the transcriptomics data (Fig. 5c). This showed that there were significant differences in the contributions of major cell lineages, with the contribution of leukocytes being higher in *FH*-deficient RCCs than in normal tissue and other subtypes of RCCs. In line with our data, this contribution was also lower in *SDH*-deficient RCCs (Fig. 5c). Finally, we also found that *FH*-deficient tumour tissue exhibited increased levels of interleukin-6 (IL-6), but not IL-10, confirming the presence of an inflammatory milieu in these tumour tissues (Fig. 5d). These results suggest that the immune phenotype associated with loss of *Fh1* that we observed in mice may be relevant to human tissue that lacks *FH*.

**Fig. 5: *FH*-deficient tumour tissue is characterized by an inflammatory signature.**




**a**, GSEA of the gene expression profile of HLRCC tumour versus normal tissue. **b**, mRNA expression of a panel of ISGs in HLRCC tumour versus normal tissue measured by qRT–PCR; N, normal (healthy individuals); T, tumour (patients with HLRCC). $n = 3$ serum samples for each N and T, in 3 technical replicates each. **c**, Cellular composition of the bulk RNA-sequencing datasets from *FH*-deficient RCCs, *SDH*-deficient RCCs and common RCC subtypes from TCGA (as described in the Methods). ccRCC, clear cell RCC; pRCC, papillary RCC; chRCC, chromophobe RCC. **d**, Levels of IL-6 and IL-10 in HLRCC tumour versus normal tissue measured by enzyme-linked immunosorbent assay (ELISA). *n*

= 5 and 20 samples for normal tissue (N) and *FH*-deficient HLRCC tumour tissue (T), respectively. **e**, Schematic of the pathway. Loss of *Fh1* in the mouse kidney results in profound metabolic adaptations and mitochondrial network remodelling mediated by fumarate within the cell. Increased levels of fumarate lead to the release of mtDNA in the cytosol through an SNX9-dependent MDV-mediated mechanism, which results in the activation of the cGAS–STING pathway as well as RIG-I, upregulation of genes involved in the innate immune response and chronic low-grade inflammation. Data are mean ± s.e.m. **b**, Unpaired two-tailed *t*-test; **c**, Wilcoxon rank sum test, $P < 0.05$ (Benjamini–Hochberg adjustment); **d**, unpaired two-tailed Mann–Whitney test.

Source data

# Discussion

In this work, we have shown that fumarate results in a remodelling of the mitochondrial network, the release of mtDNA into the cytosol and the activation of the innate immune response (Fig. 5e). Although this response emerges quickly, it persists over time, and is also seen in patients with HLRCC, suggesting that there is a direct interplay between mtDNA release, chronic inflammation and tumorigenesis in HLRCC.

Our results reveal that in addition to cGAS, RIG-I, which can recognize short viral double-stranded RNA and DNA[26] and is linked to the STING sensing pathway[27], also has a role. However, the silencing of *Mavs* had no effect on TBK1–IRF3 signalling, suggesting that RIG-I has a non-canonical function.

Although VDAC1 oligomerization[20], the opening of the mitochondrial permeability transition pore[28] and BAX and BAK pores[19] have been proposed to allow mtDNA release, we found that mtDNA is transported to the cytosol through SNX9-dependent MDVs, an alternative mechanism that preserves membrane integrity (Fig. 5e).

Overall, this study expands our understanding of how mitochondria can affect innate immune responses and provides a basis for further

investigations of metabolite-driven immunopathology.

# Methods

## Mice

Mice were of mixed genetic background C57BL/6 and 129/SvJ. Mice were bred and maintained under specific-pathogen-free conditions at the Breeding Unit (BRU) at the CRUK Cambridge Institute. *Fh1*$^{fl/fl}$ (ref. [4]) and *R26-Cre-ERT2* (ref. [5]) mice were gifts from E. Gottlieb and D. Winton, respectively. Experimental mice were homozygous for the conditional LoxP-exon3/4-LoxP *Fh1* allele and expressed the Cre recombinase–ERT2 fusion under control of the *Rosa26* promoter (*Fh1*$^{fl/fl}$*; R26 CreERT2/CreERT2*). Littermate controls lacked the LoxP-exon3/4-LoxP allele but also expressed the *CreERT2* allele under the control of the *Rosa26* promoter (*Fh1*$^{+/+}$*; R26 CreERT2/CreERT2*). Control mice were induced and euthanized at the same time as their experimental littermates. In vivo experiments (tamoxifen induction) were performed under specific-pathogen-free conditions at the BRU at the CRUK Cambridge Institute. All mouse experiments were performed in individually ventilated cages under the Animals (Scientific Procedures) Act 1986 (project licence P8A516814). The experiments were not randomized, and investigators were not blinded to treatment status during experiments and outcome assessment. Mice of both sexes were used and the sample size was initially set at 10 per group on the basis of previous experiments to account for potential losses during experiments and allow statistical analysis.

## Tamoxifen and H-151 preparation and treatment

In brief, 5 ml ethanol was added slowly to 1 g of tamoxifen (Sigma-Aldrich) in a 50-ml Falcon tube. The tamoxifen and ethanol mix was sonicated at 40% amplitude in 20-s pulses until the tamoxifen was completely dissolved. Then, 50 ml of corn oil (Sigma-Aldrich) pre-heated at 50 °C was added immediately to the tamoxifen and ethanol mix to obtain a 20 mg ml$^{-1}$ tamoxifen stock solution. The tube was then vortexed and incubated for up to 12 h in an orbital shaker at 50 °C to ensure adequate solubilization.

Aliquots of 5 ml were then stored indefinitely at −20 °C. H-151 (InvivoGen) was diluted in DMSO to obtain 20 mg $ml^{-1}$ or 4 mg $ml^{-1}$, and was further diluted in sterile phosphate-buffered saline (PBS) to obtain a 5% DMSO concentration before injection into mice. Aliquots of 2 ml were then stored at −20 °C. Before injection, the mixture was allowed to warm up to room temperature. Age-matched mice (between 10 and 12 weeks old) were used in all experiments. Each mouse received three doses of 2 mg tamoxifen by intraperitoneal injection. Tamoxifen was administered every other day to allow mice to recover between doses. Mice received a daily dose of 7 mg per kg or 1.4 mg per kg H-151 by intraperitoneal injection.

## Tissue collection

Mice were killed by cervical dislocation and the kidneys were speedily collected and processed for further analysis using electron microscopy, gene expression, mass spectrometry or staining (see below for specific applications).

## Electron microscopy

Tissue samples were fixed overnight at 4 °C in 2.5% glutaraldehyde (Sigma-Aldrich) and 2% paraformaldehyde (PFA; Polysciences) in 0.1 mol $l^{-1}$ sodium cacodylate buffer pH 7.4. The samples were post-fixed with 1% osmium tetroxide 1% (EMS) in 0.1 mol $l^{-1}$ sodium cacodylate buffer for 1 h at 4 °C. After three water washes, samples were dehydrated in a graded ethanol series and embedded in epoxy resin (Sigma-Aldrich). Ultrathin sections (60–70 nm) were obtained with an Ultrotome V (LKB) ultramicrotome, counterstained with uranyl acetate and lead citrate and viewed with a Tecnai $G^2$ (FEI) transmission electron microscope operating at 100 kV. Images were captured with a Veleta (Olympus Soft Imaging System) digital camera.

## Tissue processing for gene expression analysis

For gene expression analyses, kidneys were snap-frozen in liquid nitrogen and stored at −80 °C until processing for RNA extraction. In brief, approximately 10–20-mg chunks of snap-frozen kidney (either mouse or

human tissue) were placed in PreCellys tubes (Stretton Scientific Ltd) containing 600 μl ice-cold RNALater solution (Qiagen AllPrep DNA/RNA) for homogenization at 5,500 rpm for 2 min and spun at maximum speed for 10 min in a refrigerated centrifuge to pellet cell debris. The supernatant was used for total RNA isolation using the AllPrep DNA/RNA Kit (Qiagen) following the manufacturer's instructions. For RNA sequencing, RNA was resuspended in water and the concentration was determined using a Qubit 2.0 fluorimeter (Life Technologies) following the manufacturer's instructions. RNA quality was determined using the RNA 6000 Nano Kit (Agilent) and Agilent 2100 system Bioanalyzer system following the manufacturer's instructions.

## RNA sequencing

RNA sequencing was performed at the CRUK Cambridge Institute Genomics Core Facility using the Illumina Truseq stranded mRNA (HT) kit for RNA-sequencing polyA capture on a Illumina HiSeq4000 sequencer.

Reads were mapped to the mouse reference genome GRCm38 with the STAR (v.2.6.0c) aligner[29]. Low-quality reads (mapping quality < 20) as well as known adapters and artefacts were filtered out using Cutadapt (v.1.10.0). Read counting was performed using the Bioconductor package Rsubread (v.1.28.1) (https://github.com/LTLA/csawUsersGuide) and gene annotations from GENCODE (release M17). Differential expression analysis was carried out with DESeq2 (v.1.18.1)[30]. The conditions were contrasted against the wild-type samples. Genes were identified as differentially expressed with a false discovery rate (FDR) cut-off of 0.01 and an absolute value of $\log_2$FC ($\log_2$ of the fold change) > 0.58.

## LC–MS for untargeted metabolomics analysis of mouse tissues

Snap-frozen tissue specimens were cut and weighed into PreCellys tubes (Stretton Scientific). An exact volume of extraction solution (30% acetonitrile, 50% methanol and 20% water) was added to obtain 40 mg specimen per ml of extraction solution. Tissue samples were lysed using a PreCellys 24 homogenizer (Stretton Scientific) at 5,500 rpm for 2 min. The suspension was mixed and incubated for 1 h at −20 °C, followed by 15 min

at 4 °C in a thermomixer (Eppendorf), followed by centrifugation (16,000*g*, 15 min at 4 °C). The supernatant was collected and transferred into autosampler glass vials, which were stored at −80 °C until further analysis. To avoid bias due to machine drift, samples were randomized and processed blindly. A Dionex U3000 UHPLC system coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific) was used to perform the LC–MS analysis. Five microlitres of sample was injected onto a Sequant ZIC-pHILIC column (150 × 2.1 mm, 5 µm) and guard column (20 × 2.1 mm, 5 µm) (both Merck Millipore) for the chromatographic separation, as previously described[13]. The column oven temperature was maintained at 45 °C. The mobile phase was composed of 20 mM ammonium carbonate and 0.1% ammonium hydroxide in water (mobile phase A), and acetonitrile (mobile phase B). The flow rate was set at 0.2 ml per min with the following gradient: 80% B for 2 min, linear decrease to 20% of B 15 min. Both solvents were then brought back to the initial conditions and maintained for 8 min. The mass spectrometer was operated in full MS (50–750 *m*/*z*) and polarity switching mode. The acquired spectra were analysed using XCalibur Qual Browser and XCalibur Quan Browser software (Thermo Fisher Scientific) by referencing to an internal library of compounds.

## Metabolite analysis

LC–MS analysis was performed on a Q Exactive Orbitrap mass spectrometer coupled to a Dionex UltiMate 3000 Rapid Separation LC system (Thermo Fisher Scientific). The LC system was fitted with either a SeQuant Zic-HILIC column (column A, 150 mm × 4.6 mm, 5 µm), or a SeQuant Zic-pHILIC (column B, 150 mm × 2.1 mm, 5 µm) with the corresponding guard columns (20 mm × 2.1 mm, 5 µm) (both from Merck Millipore). With column A, the mobile phase was composed of 0.1% aqueous formic acid (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The flow rate was set at 300 µl per min and the gradient was as follows: 0–5 min 80% B, 5–15 min 30% B, 15–20 min 10% B, 20–21 min 80% B, hold at 80% B for 9 min. For column B, the mobile phase was composed of 20 mM ammonium carbonate and 0.1% ammonium hydroxide in water (solvent C), and acetonitrile (solvent D). The flow rate was set at 180 µl per min with the following gradient: 0 min 70% D, 1 min 70% D, 16 min 38% D, 16.5 min 70% D, hold at 70% D for 8.5 min. The mass

spectrometer was operated in full MS and polarity switching mode. Samples were randomized to avoid machine drift, and were blinded to the operator. The acquired spectra were analysed using XCalibur Qual Browser and XCalibur Quan Browser software (Thermo Fisher Scientific) by referencing to an internal library of compounds. Calibration curves were generated using synthetic standards of the indicated metabolites.

## Tissue staining

For H&E staining, kidney tissue was fixed in a 10% neutral buffered formalin (NBF) solution for 24 h at room temperature. The tissue was then transferred to 70% ethanol and processed for paraffin embedding. Embedding, sectioning and H&E staining were performed at the Histopathology facility at CRUK Cambridge Institute. Slides were scanned using an Axioscan Z1 and analysed using Zen Blue.

For immunohistochemistry, paraffin sections were dewaxed in two xylene baths (5 min each) and rehydrated through a series of graded ethanol solutions (100%, 90% and 70%) (5 min each). Sections were placed in distilled water before heat-induced epitope retrieval (HIER) was performed. In brief, slides were immersed in IHC-Tek Epitope Retrieval Solution (IHC World) and placed in the IHC-Tek Epitope Retrieval Steamer Set (IHC World) for 45 min to unmask the antigens using the steaming method. Once completely cooled, the slides were gently rinsed with Tris-buffered saline containing 0.1% Tween-20 (TBST) wash buffer and endogenous peroxidase activity was blocked by incubating the slides with Peroxidase Block (Dako, Agilent) for 5 min. Slides were rinsed with TBST and placed in a fresh TBST bath (5 min) before incubating for 5 min with Protein Block (Dako, Agilent). Slides were then incubated for 2 h with the primary antibody at an appropriated dilution in Antibody Diluent (Dako, Agilent). Antibodies used: anti-CD14 (1:500, anti-rabbit, PA5-95334, Invitrogen, Thermo Fisher Scientific) and anti-2SC (1:500, anti-rabbit, crb2005017, Cambridge Research Biochemicals). Slides were rinsed with TBST and placed in three fresh TBST baths (5 min each) before incubating them with Labelled Polymer-HRP Anti-Rabbit (Dako, Agilent) for 2 h. After incubation, the slides were again rinsed with TBST and placed in three fresh TBST baths (5 min each). For the detection, the EnVision+ System-HRP(DAB) kit

containing DAB chromogen concentrate (Dako, Agilent) was used for 10 min. The slides were gently rinsed with distilled water and counterstained in Mayer's Hematoxylin Solution (Sigma-Aldrich, Merck) for 10 min, followed by rinsing with tap water for 2 min. Slides were dehydrated through graded ethanol washes (70%, 90%, 100%) (5 min each) before mounting using Histomount (Thermo Fisher Scientific).

## Tissue lysates for immunoblotting

Tissue lysates were prepared in RIPA (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.5% Na-deoxycholate, 0.1% SDS in 1× PBS) buffer supplemented with phosphatases and protease inhibitor cocktails (Roche) as follows: approximately 20–30 mg of frozen tissue was homogenized in CK14/2ml PreCellys tubes for 2 min at 5,500 rpm in ice-cold RIPA buffer. Lysates were clarified by centrifugation at maximum speed for 20 min in a benchtop centrifuge (Eppendorf) at 4 °C. After clarification, the supernatant was transferred to a clean pre-chilled tube and the protein concentration was determined using the Pierce BCA Protein Assay Dye (Thermo Fisher Scientific) following the manufacturer's instructions.

## Genomic DNA amplification

Verification of the excision of exons 3 and 4 after tamoxifen recombination was performed by PCR on genomic DNA extracted from the tissues. In brief, genomic DNA was extracted using the Qiagen AllPrep DNA/RNA Kit following the manufacturer's instruction. PCR was performed on an Applied Biosystems Verity Thermo Cycler. Primers[12] were obtained from Sigma-Aldrich. PCR product was loaded on a 2% agarose gel and visualized with SybrGreen.

## Bioinformatics processing of data

The results of the differential expression analysis were visualized using volcano plots, representing the $\log_2$ of the fold change between the conditions and the adjusted *P* values obtained. GSEA was performed using the GSEA software from Broad Institute. Pre-ranked GSEA with default

settings was run to determine the enriched pathways in the different conditions versus controls, using DESeq2 metrics to rank the genes, and the Hallmark and KEGG pathway collections. The results are presented in volcano plots showing the NES and the FDR-adjusted *P* values obtained.

## Estimates of kidney tumour composition by deconvolution of transcriptomics data

### Bulk RNA-sequencing datasets

Bulk transcriptomic data for clear cell RCCs, papillary RCCs and chromophobe RCCs were downloaded from TCGA[31]. This dataset was further annotated for samples with germline *FH* mutational status[32]. Transcriptomic data for additional *FH*-deficient tumours sampled from primary and metastatic sites were also downloaded[33]. *FH*-deficient RCCs were defined as only those with a known germline mutation, and derived unequivocally from the primary kidney cancer. Samples that might have been derived from metastatic sites[33] were not classified as *FH*-deficient RCCs, as the microenvironment in metastases will differ from the primary site. In total, six tumours were unequivocally derived from primary tumour, and these were added to the four *FH*-deficient RCCs sequenced as part of the TCGA consortium. Finally, we added bulk transcriptomes from *SDH*-deficient RCCs[34].

### Single-cell-sequencing reference data

Annotated single cells derived from adult RCCs were used as a comparative reference[35]. For the mouse data, we used a mouse single-cell RNA-sequencing reference dataset[36] as reference. Bulk RNA counts are from GSE183745. Methods as above.

### Deconvolution method

We collated the reference single-cell RNA-sequencing data according to their major lineage (such as leukocytes, myeloid or endothelial), rather than using the finely annotated subtypes. The collation to major lineage types

helped to alleviate batch effects and the difficulty in discerning the presence of subtly different cell subtypes from bulk RNA data. This single-cell reference was used as input alongside the bulk RNA-sequencing data above into a Bayesian deconvolution method to allow the cellular composition of the bulk data to be estimated[37]. We compared the cellular composition of the above bulk RNA-sequencing datasets (*FH*-deficient RCCs, *SDH*-deficient RCCs and common RCC subtypes from TCGA as described above).

## Cell culture and treatment

All cell lines were grown in high-glucose (4.5 g $l^{-1}$) Dulbecco's modified Eagle's medium (DMEM) containing 1 mM sodium pyruvate and 2 mM l-glutamine (Gibco, Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS) (Gibco, Thermo Fisher Scientific). Cells were tested for mycoplasma contamination using the Lookout Mycoplasma PCR detection kit (Sigma-Aldrich). To generate $FH^{+/+\rho 0}$ cell lines, $FH^{+/+}$ cells were treated with 50 ng $ml^{-1}$ ethidium bromide for 25 days. Clones were isolated after treatment and grown in the same medium supplemented with 50 μg $ml^{-1}$ uridine (Sigma-Aldrich). MMF (Sigma-Aldrich) and dimethylsuccinate (DMS) (Sigma-Aldrich) powder were resuspended in DMSO (Thermo Fisher Scientific) at 500 mM and added to the medium at the indicated concentration for the indicated time. For long treatments, the medium was changed every two days. Cells were transfected with Lipofectamine RNA iMAX (Invitrogen) following the manufacturer's instructions, with the siRNA (Dharmacon) at a final concentration of 20 nM.

### siRNA oligonucleotides

Non-targeting siRNA scramble (ON-TARGETplus SMARTpool, D-0001810-10-20), *Bak1* (ON-TARGETplus SMARTpool, L-042978-00-0005), *Bax* (ON-TARGETplus SMARTpool, L-061976-00-0005), *Ddx58* (*Rigi*) (ON-TARGETplus SMARTpool, L-065328-00-0005), *Drp1* (*Dnm1l*) (ON-TARGETplus SMARTpool, L-054815-01-0005), *Ifih1* (*Mda5*) (ON-TARGETplus SMARTpool, L-048303-00-0005), *Mavs* (ON-TARGETplus SMARTpool, L-053767-00-0005), *Mb21d1* (*Cgas*) (ON-TARGETplus SMARTpool, L-055608-1-0005), *Mfn1* (ON-TARGETplus SMARTpool, L-

065399-01-0005), *Mfn2* (ON-TARGETplus SMARTpool, L-046303-00-0005), *Opa1* (ON-TARGETplus SMARTpool, L-042427-01-0005), *Rab9* (L-040861-01-0005), *Snx9* (ON-TARGETplus SMARTpool, L-057505-01-0005), *Tmem173* (*Sting1*) (ON-TARGETplus SMARTpool, L-055528-00-0005) and *Vdac1* (ON-TARGETplus SMARTpool, L-047345-00-0005).

## Plasmids and constructs

pEGFP and pEGFP:NDI1 were obtained from the laboratory of E. Dufour[38] and were stably expressed in $cFh1^{-/-CL1}$ cells. Lentiviral particles were generated in HEK293T cells by co-transfection of the target vector together with packaging psPAX2 (Addgene, 12260) and envelope pMD2.G (Addgene, 12259) vectors. Twenty-four hours after transduction, cells were selected for puromycin resistance. The *Fh1* reconstituted cell line ($Fh1^{-/-CL1}$*+pFH-GFP*) was generated as previously described[16]. The cytoplasmic *Fh1* cell line ($Fh1^{-/-CL1}$*+cytoFh1-GFP*) was generated by transfecting a previously generated cytoFh1-GFP plasmid[39] into cells seeded into a six-well plate using Lipofectamine 2000 (Thermo Fisher Scientific, 11668027) using 2.5 µg plasmid DNA and 9 µl Lipofectamine 2000 reagent per well. Selection was performed by incubating cells in the presence of 2 mg $ml^{-1}$ G418 (Gibco, 10131027) for two weeks.

## Antibodies

For immunoblots: goat polyclonal anti-FH/fumarase (ab113963) and anti-β-actin (AC-15) (ab6276), rabbit polyclonal anti-GRP75 (ab2799), anti-MFN1 (ab126575), anti-phospho-IRF3S386 (ab76493) and mouse monoclonal anti-VDAC1 (ab14734) antibodies were purchased from Abcam. Mouse monoclonal anti-actin (A2228) and anti-vinculin (V4505) antibodies were purchased from Sigma-Aldrich. Mouse monoclonal anti-DRP1 (611113) and anti-OPA1 (612607) antibodies were obtained from BD Transduction Laboratories. Rabbit polyclonal anti-MFN2 (11925), anti-BAK (12105), anti-BAX (2772), anti-cGAS (31659S), anti-IRF3 (4302S), anti-STAT1 (9172S), anti-pSTAT1 Tyr701 (9167S), anti-TBK1/NAK (3013S), anti-phospho-TBK1/NAK Ser172 (5483S), anti-RAB9A (5118S), anti-RIG-I (3743S) and anti-STING (50494S) antibodies were obtained from Cell Signaling Technology. Rabbit polyclonal anti-SNX9 (15721-1-AP) was

purchased from Proteintech. Mouse monoclonal anti-cytochrome *c* (556433) antibody was purchased from BD Pharmingen. Rabbit polyclonal anti-mtTFAM (GTX103231) antibody was obtained from GeneTex.

For immunofluorescence: rabbit polyclonal anti-mtTFAM (GTX103231) antibody from GeneTex was used and mouse monoclonal anti-DNA (CBL186) antibody was purchased from Millipore. Rabbit polyclonal anti-TOM20 (ab232589), mouse monoclonal anti-TOM20 (ab56783), mouse anti-PDH (ab110333), rabbit anti-pTBK1 (ab109272) and mouse anti-TOM22 (ab57523) antibodies were purchased from Abcam. Mouse anti-cytochrome *c* (556432) antibody was purchased from BD Pharmingen, mouse anti-GM130 (610822) antibody was purchased from BD BioSciences and rabbit polyclonal anti-cGAS (D3080) antibody was purchased from Cell Signaling Technology. Donkey anti-mouse, goat anti-mouse IgG1, goat anti-mouse IgG2a, goat anti-mouse IgGM and goat anti-rabbit Alexa Fluor 488, 565, 594 or 647 were used as secondary antibodies (Invitrogen).

## Immunofluorescence

Immunofluorescence were performed as previously described[40]. In brief, cells were fixed in 5% PFA in PBS at 37 °C for 15 min, then washed three times with PBS, followed by quenching with 50 mM ammonium chloride in PBS. After three washes in PBS, cells were permeabilized in 0.1% Triton X-100 in PBS for 10 min, followed by three washes in PBS. Cells were then blocked with 10% FBS in PBS, followed by incubation with primary antibodies in 5% FBS in PBS, for 2 h at room temperature. Cells were then washed three times in 5% FBS in PBS and incubated with the corresponding secondary antibody (dilution 1:1,000) prepared in 5% FBS in PBS. After three washes in PBS, coverslips were briefly rinsed in water and mounted onto slides using Dako fluorescence mounting medium (Dako) or in ProLong Diamond Antifade Mountant (P369621) purchased from Thermo Fisher Scientific.

## Confocal microscopy

Stained cells were imaged using a 100× objective lense (NA 1.4) on a Nikon Eclipse TiE inverted microscope with appropriate lasers using an Andor

Dragonfly 500 spinning disk system, equipped with a Zyla 4.2 PLUS sCMOS camera (Andor), coupled with Fusion software. For mitochondrial morphology analysis, nine stacks of 0.2 μm each were acquired using the 100× objective. Images acquired in the same conditions of laser intensity and exposure time from the same experiment were then compiled by 'max projection' and mitochondrial morphology was analysed and presented as intermediate, elongated, swollen, swollen-elongated or fragmented. Mitochondrial morphology was also quantified in a semi-automated and unbiased manner using a modified version of MitoMAPR (ref. [41]). In brief, regions of interest (ROIs) were pre-processed using background subtraction (rolling ball 50), followed by 1× unsharp mask and 1× smooth filter functions in ImageJ. The images were then thresholded and converted to binary. To separate closely connected mitochondrial particles, we used the built-in watershed algorithm in ImageJ (ref. [42]). The images were batch processed using this workflow to calculate the mitochondrial area ($\mu m^2$) of each mitochondrial particle and the number of mitochondrial particles within a ROI using the 'analyse particle' function of ImageJ. For the quantification of cytosolic DNA foci, the number of DNA foci outside the nucleus and the mitochondrial perimeter was counted manually. Cells were considered as 'releasing cytosolic DNA foci' when the number of DNA foci counted was higher than two per cell. The number of $TOM20^{-}PDH^{+}$ and $TOM20^{-}PDH^{+}DNA^{+}$ MDVs outside mitochondria was assessed manually. Cells were considered as 'releasing vesicles' when the number of DNA foci counted was higher than one per cell.

Representative images were processed once with the 'smooth' function in Fiji.

## Super-resolution microscopy

Super-resolution images (Fig. 4c and Extended Data Fig. 3a) were acquired with a Nikon N-SIM microscope. Nine *z*-stacks of 0.2 μm each were acquired using a SR Apo TIRF 100× 1.49 N.A. oil objective and a DU897 Ixon camera (Andor). Raw images were computationally reconstructed using the reconstruction slice system from NIS-Elements software (Nikon). Mitochondrial diameter was calculated using the Fiji software.

Super-resolution images related to $TOM20^{-}PDH^{+}TFAM^{+}$ MDVs (Extended Data Fig. 11d) were acquired with a Zeiss LSM 880 microscope with an Airyscan detector (Carl Zeiss Microscopy), using a Zeiss 63× oil lens, NA 1.4, as the primary objective. Four-colour excitation was performed with a blue diode laser for 405 nm, argon laser for 488 nm, He 543 laser for 561 nm and He 633 laser for 647 nm. The Airyscan detector was used in SR mode using all 32 pinholes, thus increasing the resolution to around 140 nm in *x*, *y* and *z* to capture the image. The image was reconstructed by pixel reassignment and deconvolution on the Zen Black platform (Carl Zeiss Microscopy). The final image was produced on ImageJ by background subtraction (rolling ball 50) and one run of the smooth filter.

## Characterization of MDVs by super-resolution microscopy

Images were acquired with a Zeiss Elyra7 equipped with Lattice $SIM^2$ (Carl Zeiss Microscopy) with a 15-phased imaging protocol for all channels (Extended Data Fig. 11e). A Zeiss 63× OIL: Plan-Apo 63×/1.4 Oil Corr WD: 0.35 was used as the primary objective to obtain representative images of individual cells. The images were then reconstructed on the Zen Black platform (Carl Zeiss Microscopy) with a two-step $SIM^2$ function. The reconstructed images after deconvolution have an approximate *x*, *y* resolution of 80 nm and *z* resolution of 200 nm. The images used for quantifying the number and dimensions of MDVs were acquired on a 40× OIL: Plan-APO 40×/1.4 Oil DIC (UV) VIS-IR objective using the SIM Apotome mode to maximize the field of view and include more than one cell at a higher resolution.

MDVs were selected on the basis of the following criteria: positive for the markers PDH, DNA, but negative for TOM20. ROIs positive for bona fide MDVs were duplicated and saved for quantification.

MDVs were batch-quantified in an unbiased manner using a MACRO script written for ImageJ. In brief, the script splits the images and selects the PDH channel, which is then converted into binary mask for quantification. Using the ‘Analyse particle’ function of Fiji, the following attributes of the MDVs were generated and stored as a .csv file: area of individual MDV ($\mu m^2$),

maximum Feret's diameter (µm), minimum Feret's diameter (µm) and circularity of individual MDV.

As can be seen from the distribution of the circularity values, most of the MDVs are of irregular shape (circularity of a true circle ($4\pi \times \{\text{area}/\text{perimeter}^2\} = 1$). Thus, we used an average of the maximum and minimum Feret's diameter to represent the dimensions of an MDV. For calculating the area or size of an individual MDV, ImageJ uses an ellipse fitting strategy as described previously[42]. We used 400 individual MDVs to generate an average area ($\mu m^2$) and average Feret's diameter distribution (µm) of MDVs.

## Measuring mitochondrial volume by TEM

To measure the surface volume density, a square test-grid was placed over three randomly selected cell areas of the cell of interest. The number of test-points ($P$) inside an organelle was calculated, and the number of intersections ($I$) between the mitochondria and the test-grid was used to measure the surface density of the organelle membrane ($Sv = 2\sum I/d\sum P$, in $\mu m^2/\mu m^3$, where $d$ is the real distance between the test-lines in the defined image). The absolute surface area was calculated by multiplying the surface volume density of the organelle of interest by the organelle volume. In brief, vertical serial 1-µm sections were made through the cell of interest, and then the section where the nucleus diameter was maximal was re-embedded. Next, serial ultrathin vertical sections were obtained. The section where the width of the nucleus was maximal was considered as the central section, and the vertical axis was generated through the nuclear centre. Then the absolute organelle volumes were measured using the discretized version of vertical rotator. To test whether differences were significant (that is, $P < 0.05$), Student's $t$-tests were used. A difference was considered significant when $P < 0.05$.

## Cell fractionation

Cells were washed three times with cold PBS and scraped over ice in heavy membranes isolation buffer (HMIB) (220 mM mannitol, 70 mM sucrose, 10 mM HEPES pH 7.5, 1 mM EGTA and Roche complete protease inhibitor

cocktail). Cells were broken with manual tissue grind (Kimble). Samples were centrifuged at 800*g* for 10 min at 4 °C. Post-nuclear supernatants were then centrifuged at 2,300*g* for 10 min at 4 °C to pellet the heavy mitochondrial fraction (crude mitochondria). Pellets were washed and re-centrifuged at 4 °C for a further 15 min at 9,000*g*. The post-heavy membrane supernatants were then centrifuged at 100,000*g* at 4 °C for 60 min to obtain the cytosolic fraction. Total extracts, heavy membranes and cytosolic fractions were extracted with HMIB 1% Triton X-100, normalized for protein content using a BioRad protein assay (BioRad) and processed for SDS–PAGE and immunoblotting.

## Isolation of cytosolic fractions and ddPCR

Cytosolic fractions were obtained by extracting $10^5$ cells with 25 μg $ml^{-1}$ digitonin in isolation buffer (150 mM NaCl and 50 mM HEPES) for 10 min on ice. Samples were then centrifuged at 800*g* for 5 min at 4 °C. Supernatants were then further centrifuged at 25,300*g* for 10 min at 4 °C and pellets were discarded. Samples were normalized for protein concentration using a BioRad protein assay (BioRad). DNA was extracted from equal amounts of cytosolic fractions using a DNA extraction kit (Qiagen) following the manufacturer's instructions. mtDNA copy number was assessed using ddPCR (Biorad) using specific probes directed against mtCo3, mtND1 and mtD-loop.

Primer sequences (IDT):

mtND1_for: 5′-GAGCCTCAAACTCCAAATACTCACT-3′

mtND1_rev: 5′-GAACTGATAAAAGGATAATAGCTATGGTTACTTCA-3′

mtCo3_for: 5′-CCTCGTACCAACACATGATCTAGG-3′

mtCo3_rev: 5′-AGTGGGACTTCTAGAGGGTTAAGTG-3′

mtDloop_for: 5′-AATCTACCATCCTCCGTGAAACC-3′

mtDloop_rev: 5′-TCAGTTTAGCTACCCCCAAGTTTAA-3′

Fluorescent probe sequences (PrimeTime 5′HEX/FAM/3′ BHQ-1; IDT):

FAM-mtDloop1: 5′-ACCAATGCCCCTCTTCTCGCTCC-3′

HEX-mtND1: 5′-CCGTAGCCCAAACAAT-3′

HEX-mtCo3: 5′-ACCTCCAACAGGAATTTCA-3′.

## Extraction of RNA from cells

Cells were homogenized using QIAshredder (Sigma-Aldrich) and RNA was extracted using the Qiagen AllPrep DNA/RNA Kit according to the manufacturer's instructions. RNA concentration was determined using a Nanodrop (ND1000). Reverse transcription was performed using 800 ng total RNA with the Applied Biosystems High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific).

## qRT–PCR

qRT–PCR was performed on 5 ng of cDNA using a final concentration of 2 μM each of forward and reverse primers (see primers list) using Quantitect Syber Green Master Mix (Qiagen) on an Applied Biosystems StepOne Plus or QuantStudio5 real-time PCR system. Experiments were analysed using the ΔΔCt method. Statistical analysis was performed using Prism 7 software. Samples were normalized relative to the expression of an endogenous control gene (*Rplp0*; see primers list for sequences). Amplification primers were designed using Primer 3.1 to span one intron and purchased from Sigma-Aldrich, and are listed in Supplementary Tables 4 and 5.

## OCR measurements (Seahorse)

The OCR was measured using the real-time flux analyser XF-24e (Seahorse Bioscience) as previously described[12]. In brief, $4 \times 10^4$ $iFh1^{fl/flCL29}$ or $iFh1^{fl/flCL33}$ cells were treated with either vehicle (ethanol) or 4-OHT for 48 h, after which fresh medium was added. The cells were then kept in culture for 3, 10, 15 or 21 days before starting the OCR assay. For this, cells were treated with 1 μM oligomycin, 2 μM carbonyl cyanide-p-

trifluoromethoxyphenylhydrazone (FCCP), rotenone and antimycin A (both 1 μM) (all Sigma-Aldrich). At the end of the run, cells were lysed using RIPA buffer (25 mM Tris/HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate and 0.1% SDS). The protein content of each well was measured using a BCA kit (Pierce) following the manufacturer's instructions. OCR was normalized to the total protein content as indicated.

## Oxygen consumption experiments (Oroboros)

Cells were seeded on 10-cm dishes to reach more than 90% confluency on the day of experiments. Cells were rinsed with PBS and trypsinized followed by a short spin at 300*g* for 3 min at room temperature. The cell pellet was resuspended in complete DMEM devoid of FBS, and the cell density was determined using a Countless 3 automated cell counter (Thermo Fisher Scientific). Oxygen consumption on intact cells was monitored polarographically with an Oxygraph-2k (Oroboros Instruments) at 37 °C in a 2-ml chamber with complete DMEM without FBS, calibrated daily using air-saturated reaction medium and a correction factor of 0.89. Cell suspension was added to the chamber and allowed to equilibrate (2–3 min) before closing the chamber and starting the record. After a stable basal respiration was achieved (5 min), oligomycin (5 μM) was added to assess ATP-dependent respiration (5 min), followed by a titration with the uncoupler FCCP at 0.5-μM pulses to determine maximal uncoupled respiration (180 s each). Finally, rotenone (0.5 μM) was added to inhibit Complex I-linked respiration. Respiration rates were calculated as the average value over a 30-s window in DatLab 7.4 (Oroboros Instruments) and are expressed in pmol $O_2$ per second per $10^6$ cells.

## SDS–PAGE and immunoblotting

Cells were lysed in RIPA buffer (20 mM Tris pH 8.0, 150 mM NaCl, 0.1% SDS, 1% deoxycholic acid, 1% NP-40 and complete protease inhibitor cocktail (Roche)). Alternatively, PathScan(R) Sandwich ELISA lysis buffer (Cell Signaling Technology) supplemented with complete protease inhibitor cocktail and phospho-stop (Sigma-Aldrich) was used to extract samples to resolve phosphoproteins. Samples were normalized for protein concentration using a BioRad protein assay (BioRad). Proteins were resolved by SDS–

PAGE and transferred to nitrocellulose membranes (0.2 μm pore size, GE Healthcare) or PVDF membrane (0.2 μm pore size, GE Healthcare).

For immunoblot analysis, membranes were blocked with 2–5% non-fat milk or 2% bovine serum albumin (BSA) in PBS for 60 min at room temperature. Membranes were incubated with primary antibody at the appropriate dilution (in 2–5% non-fat milk or 2% BSA in 0.05% Tween-20 in PBS) at 4 °C overnight. Membranes were washed in 0.05% Tween-20 in PBS three times for 15 min and incubated with appropriate secondary antibodies (1:3,000 in 2–5% milk or 2% BSA in 0.05% Tween-20 in PBS). Membranes were treated with Western Lightning Plus ECL (Perkin Elmer), and exposed to chemiluminescence either on films (PROTEC) or on a digital ECL machine (Amersham) for image quantification. Uncropped immunoblots are presented in Supplementary Fig. 10.

## ELISA

Serum levels of IL-6, IL-10 and IFNβ were measured by ELISA (Biotechne) according to the manufacturer's instructions. Ninety-six-well Maxisorp (Thermo Fisher Scientific) plates were coated with 100 μl of capture antibody diluted with PBS (2 μg $ml^{-1}$ IL-6, 2 μg $ml^{-1}$ IL-10 and 4 μg $ml^{-1}$ IFNβ) and sealed overnight at room temperature. Plates were then washed four times with 0.05% PBS-Tween-20 (PBS-T). After blotting, plates were blocked with 300 μl of 1% BSA in PBS for 1 h at room temperature. After repeated washing, either 50 μl (IL-6 and IL-10) or 100 μl (IFNβ) of samples was added in duplicate; 1:10 dilutions in 1% BSA were performed for all samples. Standard concentrations ranged from 600 pg $ml^{-1}$ to 9.38 pg $ml^{-1}$ for IL-6; 2,000 pg $ml^{-1}$ to 31.3 pg $ml^{-1}$ for IL-10; and 500 pg $ml^{-1}$ to 7.81 pg $ml^{-1}$ for IFNβ. Plates were sealed and incubated at room temperature for 2 h. The washing step was repeated and either 50 μl (IL-10) or 100 μl (IL-6 and IFNβ) of detection antibody diluted in PBS was added (50 ng $ml^{-1}$ IL-6, 50 ng $ml^{-1}$ IL-10 and 250 ng $ml^{-1}$ IFNβ). The plate was sealed for 2 h at room temperature.

Plates were then washed four times and 100 μl of 1:40 streptavidin–HRP diluted in 1% BSA was added per well. Plates were sealed and incubated for 20 min at room temperature, followed by four washes. Tetramethylbenzidine

(TMB) substrate solution (100 µl) was added for up to 15 min, followed by 50 µl of 1 mol $l^{-1}$ $H_2SO_4$ to stop the reaction. Optical density was read immediately in a plate reader (BioTek Synergy H1) at an absorbance of 450 nm.

## Number of cells used for image quantification

For image quantification, experiments are presented from three independent experiments (unless otherwise specified) with a total number of cells (*n*) analysed as: main figures: Fig. 2d–f: *n* = 173, 157, 158, 170, 160, 157, 162 for NT, vehicle, d1, d3, d6, d10, d15, respectively. Fig. 3a–c: *n* = 221, 192, 185 for vehicle, 200 µM MMF, 400 µM MMF, respectively. Fig. 4d,e: for $TOM20^{-}PDH^{+}DNA^{+}$ vesicle analysis: *n* = 166, 180, 167, 155, 170 for NT, 2d, 4d, 6d, 8d, respectively. Fig. 4e: for cytosolic DNA foci analysis: *n* = 165, 171, 162, 152, 160 for NT, 2d, 4d, 6d, 8d, respectively. Fig. 4f: *n* = 156; number of vesicles analysed: *n* = 155. Fig. 4h,i: for $TOM20^{-}PDH^{+}DNA^{+}$ vesicle analysis: *n* = 158, 173, 158 for NT, MMF 6d siScr, MMF 6 days si*Snx9*, respectively. Fig. 4i: for cytosolic DNA foci analysis: *n* = 151, 166, 197, for NT, MMF 6d siScr, MMF 6d si*Snx9*, respectively. Extended Data: Extended Data Fig. 2g: *n* = 173, 157, 158, 170, 160, 157, 162 for NT, vehicle, d1, d3, d6, d10, d15, respectively. Extended Data Fig. 2h,i: *n* = 30, 32, 35, 38, 32, 40, 36 and number of ROIs analysed: *n* = 84, 85, 84, 85, 85, 84, 85 for NT, vehicle, d1, d3, d6, d10, d15, respectively. Extended Data Fig. 2j: number of mitochondria analysed: 11, 22, 79, 54, 46 for $cFh1^{+/+}$, $cFh1^{+/+}$ + Cre, $cFh1^{fl/fl}$, $cFh1^{-/-CL1}$, $cFh1^{-/-CL19}$, respectively. Extended Data Fig. 3b: 30 cells were analysed per condition with a number of mitochondria analysed of *n* = 429, 827, 780 for $cFh1^{fl/fl}$, $cFh1^{-/-CL1}$, $cFh1^{-/-CL19}$, respectively. Extended Data Fig. 3c,f,g: *n* = 272, 246, 306, 300 for $cFh1^{fl/fl}$, $cFh1^{-/-CL1}$, $cFh1^{-/-CL19}$, $cFh1^{-/-CL1}$ + *pFH-GFP*. Extended Data Fig. 3d: *n* = 33, 45, 67, 39 and number of ROIs analysed: *n* = 79, 71, 164, 73 for $cFh1^{fl/fl}$, $cFh1^{-/-CL1}$, $cFh1^{-/-CL19}$, $cFh1^{-/-CL1}$ + *pFH-GFP*, respectively. Extended Data Fig. 4e: *n* = 221, 192, 185 for vehicle, 200 µM MMF, 400 µM MMF, respectively. Extended Data Fig. 5a: *n* = 221, 192, 185 for vehicle, 200 µM MMF, 400 µM MMF, respectively. Extended Data Fig. 5e: *n* = 169, 188 for DMSO and MMF 8d, respectively. Extended Data Fig. 6g: *n* = 182, 156, 189, 228 for $cFh1^{fl/fl}$, $cFh1^{-/-CL1}$, $cFh1^{-/-CL1}$ + pcytoFH-GFP, $cFh1^{-/}$

$^{-CL1}$ + *pFH*-GFP, respectively. Extended Data Fig. 7b,c: $n = 152$, 153, 162, 155, 155 for vehicle, DMS 200 µM, DMS 400 µM, DMS 1 mM, DMS 5 mM, respectively. Extended Data Fig. 8d: $n = 933$, 660 for DMSO and MMF 8d, respectively. Extended Data Fig. 9a: $n = 175$, 171, 183, 168, 166, 187, 188, 186, 154, 184, 161, 180, 169, for $cFh1^{fl/fl}$, $cFh1^{-/-CL1}$ siScr, $cFh1^{-/-CL1}$ si*Snx9*, $cFh1^{-/-CL1}$ si*Vdac1*, $cFh1^{-/-CL1}$ si*Bax*/*Bak1*, $cFh1^{-/-CL1}$ si*Rab9*, $cFh1^{-/-CL1}$ si*Drp1*, $cFh1^{-/-CL1}$ si*Opa1*, $cFh1^{-/-CL1}$ si*Mfn1*, $cFh1^{-/-CL1}$ si*Mfn2*, $cFh1^{-/-CL\ 1}$ si*Cgas*, $Fh1^{-/-}$ si*Sting1*, $cFh1^{-/-CL1}$ si*Rig-I*, respectively. Extended Data Fig. 9e: $n = 31$, 32, for $cFh1^{fl/fl}$ and $cFh1^{-/-CL1}$, respectively. Extended Data Fig. 10b,c: $n = 151$, 154, 166, 159, 161, 163, for $cFh1^{fl/fl}$, $cFh1^{fl/fl}$ siScr, $cFh1^{fl/fl}$ si*Snx9*, $cFh1^{-/-CL1}$ siScr, $cFh1^{-/-CL1}$ si*Snx9*, $cFh1^{-/-CL19}$ siScr, $cFh1^{-/-CL19}$ si*Snx9*, respectively. Extended Data Fig. 10e,f: $n = 55$, 50, 53, 55, 60, 53 and number of ROIs analysed: $n = 168$, 162, 158, 348, 125, 155, 76 for $cFh1^{fl/fl}$ + siScr, $cFh1^{fl/fl}$ + si*Snx9*, $cFh1^{-/-CL1}$ + siScr, $cFh1^{-/-CL1}$+ si*Snx9*, $cFh1^{-/-CL19}$ + siScr, $cFh1^{-/-CL19}$ + si*Snx9*, respectively. Extended Data Fig. 10g,h: $n = 191$, 167, 160, 179. 190, 147, 160, for vehicle, d6 siScr, d6 si*Snx9*, d10 siScr, d10 si*Snx9*, d15 siScr, d15 si*Snx9*, respectively. Extended Data Fig. 11a–c: $n = 166$, 180, 167, 155, 170, for NT, 2d, 4d, 6d, 8d MMF, respectively. Extended Data Fig. 11d: $n = 85$. Extended Data Fig. 11f,g: $n = 408$ ROIs analysed. Extended Data Fig. 11i: $n = 189$, 155, 272, 159, 159, 144 for vehicle, 4-OHT d1, d3, d6, d10, d15, respectively. Extended Data Fig. 12b: $n = 170$, 167, 145, 136, 133, 136, 139 for $Fh1^{+/+}$ + DMSO, $Fh1^{+/+}$ + MMF 6d, $Fh1^{+/+r0}$ + DMSO $Fh1^{+/+r0}$ + MMF 1d, 3d, 6d, 8d, respectively. Extended Data Fig. 12c: $n = 151$, 166, 197 for NT, MMF 6d siScr, MMF 6d si*Snx9*, respectively. Supplementary Information: Supplementary Fig. 1f–h: $n = 166$, 159, 166, 165, 175, 181, 170 for NT, vehicle, d1, d3, d6, d10, d15 respectively. Supplementary Fig. 1i,j: $n = 35$, 42, 43, 47, 50, 48, 36 and number of ROI analysed: $n = 99$, 115, 137, 96, 127, 127, 76 for NT, vehicle, d1, d3, d6, d10, d15, respectively. Supplementary Fig. 3e: $n = 163$, 161, 163 for $Sdhb^{fl/fl}$, $Sdhb^{-/-CL5}$, $Sdhb^{-/-CL7}$, respectively. Supplementary Fig. 4c: $n = 152$, 153, 162, 155, 155 for vehicle, DMS 200 µM, DMS 400 µM, DMS 1 mM, DMS 5 mM, respectively. Supplementary Fig. 3f,g: $n = 163$, 161, 163 for $Sdhb^{fl/fl}$, $Sdhb^{-/-CL5}$, $Sdhb^{-/-CL7}$, respectively. Supplementary Fig. 4d: $n = 174$, 165, 171 for $cFh1^{fl/fl}$, $cFh1^{-/-CL1}$ + pEGFP, $cFh1^{-/-CL1}$ + pEGFP:NDI1, respectively. Supplementary Fig. 9b: $n = 408$ ROIs analysed.

### Human tissue samples

All samples were from adult individuals with confirmed *SDH*- or *FH*-deficient renal cancer, with informed consent (Research Ethics Committee approval reference 16/WS/0039).

### Statistical analysis

Errors bars in graphs represent the mean ± s.e.m. from at least three independent experiments. The number of cells analysed is provided in the Methods. Statistical significance was analysed using a two-tailed, unpaired, Student's *t*-test, Mann–Whitney test or one-way or two-way ANOVA. All statistical analyses were performed using Excel or Prism software.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All data are included within the article or the supplementary information. Full versions of all gels and blots are provided in Supplementary Fig. 10. See the Supplementary Information guide for details. Raw FastQ files for RNA-sequencing analyses are publicly available in the Gene Expression Omnibus (GEO) repository with the accession code GSE183745. All materials generated in this study are readily available from the authors. All other data supporting the findings of this study are available from the corresponding authors upon request. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising. Source data are provided with this paper.

# References


1. The Multiple Leiomyoma Consortium. Germline mutations in *FH* predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nat. Genet.* **30**, 406–410 (2002).

2. Schmidt, C., Sciacovelli, M. & Frezza, C. Fumarate hydratase in cancer: a multifaceted tumour suppressor. *Semin. Cell Dev. Biol.* **98**, 15–25 (2020).

3. Pollard, P. J. et al. Accumulation of Krebs cycle intermediates and over-expression of HIF1α in tumours which result from germline *FH* and *SDH* mutations. *Hum. Mol. Genet.* **14**, 2231–2239 (2005).

4. Pollard, P. J. et al. Targeted inactivation of *Fh1* causes proliferative renal cyst development and activation of the hypoxia pathway. *Cancer Cell* **11**, 311–319 (2007).

5. Hameyer, D. et al. Toxicity of ligand-dependent Cre recombinases and generation of a conditional Cre deleter mouse allowing mosaic recombination in peripheral tissues. *Physiol. Genomics* **31**, 32–41 (2007).

6. Zheng, L. et al. Reversed argininosuccinate lyase activity in fumarate hydratase-deficient cancer cells. *Cancer Metab.* **1**, 12 (2013).

7. Alderson, N. L. et al. *S*-(2-succinyl)cysteine: a novel chemical modification of tissue proteins by a Krebs cycle intermediate. *Arch. Biochem. Biophys.* **450**, 1–8 (2006).

8. Mathew, R., Degenhardt, K., Haramaty, L., Karp, C. M. & White, E. Immortalized mouse epithelial cell models to study the role of apoptosis in cancer. *Methods Enzymol.* **446**, 77–106 (2008).

9. Okude, H., Ori, D. & Kawai, T. Signaling through nucleic acid sensors and their roles in inflammatory diseases. *Front. Immunol.* **11**, 625833 (2021).

10. Riley, J. S. & Tait, S. W. Mitochondrial DNA in inflammation and immunity. *EMBO Rep.* **21**, e49799 (2020).

11. Yang, Y. et al. UOK 262 cell line, fumarate hydratase deficient (*FH⁻/FH⁻*) hereditary leiomyomatosis renal cell carcinoma: in vitro and in vivo model of an aberrant energy metabolic pathway in human cancer. *Cancer Genet. Cytogenet.* **196**, 45–55 (2010).

12. Frezza, C. et al. Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* **477**, 225–228 (2011).

13. Cardaci, S. et al. Pyruvate carboxylation enables growth of SDH-deficient cells by supporting aspartate biosynthesis. *Nat. Cell Biol.* **17**, 1317–1326 (2015).

14. Cho, J., Hur, J. H., Graniel, J., Benzer, S. & Walker, D. W. Expression of yeast NDI1 rescues a *Drosophila* complex I assembly defect. *PLoS ONE* **7**, e50644 (2012).

15. Tyrakis, P. A. et al. Fumarate hydratase loss causes combined respiratory chain defects. *Cell Rep.* **21**, 1036–1047 (2017).

16. Sciacovelli, M. et al. Fumarate is an epigenetic modifier that elicits epithelial-to-mesenchymal transition. *Nature* **537**, 544–547 (2016).

17. Xiao, M. et al. Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* **26**, 1326–1338 (2012).

18. Vincent, J. et al. Small molecule inhibition of cGAS reduces interferon expression in primary macrophages from autoimmune mice. *Nat. Commun.* **8**, 750 (2017).

19. McArthur, K. et al. BAK/BAX macropores facilitate mitochondrial herniation and mtDNA efflux during apoptosis. *Science* **359**, eaao6047 (2018).

20. Kim, J. et al. VDAC oligomers form mitochondrial pores to release mtDNA fragments and promote lupus-like disease. *Science* **366**, 1531–1536 (2019).

21. Tilokani, L., Nagashima, S., Paupe, V. & Prudent, J. Mitochondrial dynamics: overview of molecular mechanisms. *Essays Biochem.* **62**, 341–360 (2018).

22. West, A. P. et al. Mitochondrial DNA stress primes the antiviral innate immune response. *Nature* **520**, 553–557 (2015).

23. Sugiura, A., McLelland, G., Fon, E. A. & McBride, H. M. A new pathway for mitochondrial quality control: mitochondrial-derived vesicles. *EMBO J.* **33**, 2142–2156 (2014).

24. Sugiura, A., Mattie, S., Prudent, J. & Mcbride, H. M. Newly born peroxisomes are a hybrid of mitochondrial and ER-derived pre-peroxisomes. *Nature* **542**, 251–254 (2017).

25. Matheoud, D. et al. Parkinson's disease-related proteins PINK1 and Parkin repress mitochondrial antigen presentation. *Cell* **166**, 314–327 (2016).

26. Chiu, Y. H., MacMillan, J. B. & Chen, Z. J. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell* **138**, 576–591 (2009).

27. Zevini, A., Olagnier, D. & Hiscott, J. Crosstalk between cytoplasmic RIG-I and STING sensing pathways. *Trends Immunol.* **38**, 194–205 (2017).

28. Yu, C. H. et al. TDP-43 triggers mitochondrial DNA release via mPTP to activate cGAS/STING in ALS. *Cell* **183**, 636–649 (2020).

29. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

30. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

31. Ricketts, C. J. et al. The Cancer Genome Atlas comprehensive molecular characterization of renal cell carcinoma. *Cell Rep.* **23**, 313–326 (2018).

32. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of papillary renal-cell carcinoma. *N. Engl. J. Med.* **374**, 135–145 (2016).

33. Crooks, D. R. et al. Mitochondrial DNA alterations underlie an irreversible shift to aerobic glycolysis in fumarate hydratase-deficient renal cancer. *Sci. Signal.* **14**, eabc4436 (2021).

34. Neves, J. B. et al. Defining the origin, evolution, and immune composition of SDH-deficient renal cell carcinoma. *iScience* **25**, 105389 (2022).

35. Young, M. D. et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594–599 (2018).

36. Park, J. et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* **360**, 758–763 (2018).

37. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat. Cancer* **3**, 505–517 (2022).

38. Cannino, G. et al. Glucose modulates respiratory complex I activity in response to acute mitochondrial dysfunction. *J. Biol. Chem.* **287**, 38729–38740 (2012).

39. Johnson, T. I., Costa, A. S. H., Ferguson, A. N. & Frezza, C. Fumarate hydratase loss promotes mitotic entry in the presence of DNA damage after ionising radiation. *Cell Death Dis.* **9**, 913 (2018).

40. Nagashima, S. et al. Golgi-derived PI(4)P-containing vesicles drive late steps of mitochondrial division. *Science* **367**, 1366–1371 (2020).

41. Zhang, Y. et al. Neuronal TORC1 modulates longevity via AMPK and cell nonautonomous regulation of mitochondrial dynamics in *C. elegans*. *eLife* **8**, e49158 (2019).

42. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2009).

## Acknowledgements

This work was supported by the Medical Research Council to J.P. (MC_UU_00015/7 and MC_UU_00028/5) and C.F. (MRC_MC_UU_12022/6). V.P. was supported by the European Union's Horizon 2020 research and innovation programme (MITODYN-749926) (2017–2019). V.Z. was supported by a WWCR grant (14-0319). I.M.N.W. was supported by a grant from the Nora Baart Foundation. T.Y. was supported by the European Research Council Consolidator Award to C.F. (ERC819920). E.N. was supported by a CRUK Programme Foundation award to C.F. (C51061/A27453). J.L.M. was supported by a MRC-funded graduate student fellowship. G.C.P. is supported by the Swiss National Science Foundation (Synergia project CRSII5_180326). S.N. was the recipient of a Daiichi Sankyo Foundation of Life Science postdoctoral fellowship (2017–2019). A.F. was supported by a Rotary Global Grant. M.S.-M is supported by CRUK. We thank M. Micaroni for the quantification of the volume of mitochondria from the TEM in Extended Data Fig. 2j. We apologize for some relevant studies not being cited in the manuscript owing to space limitations.

## Funding

Open access funding provided by Universität zu Köln and MRC Mitochondrial Biology Unit, University of Cambridge.

## Author information

Author notes

1. Irene Herranz-Montoya

   Present address: Molecular Oncology Programme, Growth Factors, Nutrients and Cancer Group Centro Nacional de Investigaciones Oncológicas (CNIO), Madrid, Spain

2. Joëlle Janssen

   Present address: Human and Animal Physiology, Wageningen University and Research, Wageningen, The Netherlands

3. Inge M. N. Wortel

   Present address: Department of Data Science, Institute for Computing and Information Sciences, Radboud University, Nijmegen, The Netherlands

4. Marc Segarra-Mondejar, Ming Yang & Christian Frezza

   Present address: CECAD Research Centre, University of Cologne, Cologne, Germany

5. Ana S. H. Costa

   Present address: Matterworks, Somerville, MA, USA

6. Laura Tronci

   Present address: Cogentech SRL Benefit Corporation, Milan, Italy

7. Dóra Bihary

Present address: VIB KU Leuven Center for Cancer Biology, Leuven, Belgium

8. Shun Nagashima

   Present address: Laboratory of Regenerative Medicine, School of Life Sciences, Tokyo University of Pharmacy and Life Sciences, Tokyo, Japan

9. These authors contributed equally: Vincent Zecchini, Vincent Paupe

10. These authors jointly supervised this work: Julien Prudent, Christian Frezza

## Authors and Affiliations

1. Medical Research Council Cancer Unit, University of Cambridge, Cambridge, UK

   Vincent Zecchini, Irene Herranz-Montoya, Joëlle Janssen, Inge M. N. Wortel, Ashley Ferguson, Marc Segarra-Mondejar, Ana S. H. Costa, Laura Tronci, Timothy Young, Efterpi Nikitopoulou, Ming Yang, Dóra Bihary, Alyson Speed, Shamith Samarajiwa & Christian Frezza

2. Medical Research Council Mitochondrial Biology Unit, University of Cambridge, Cambridge, UK

   Vincent Paupe, Jordan L. Morris, Suvagata Roy Chowdury, Gonçalo C. Pereira, Shun Nagashima, Mark Johnson & Julien Prudent

3. Department of Biology, University of Padova, Padova, Italy

   Federico Caicci

4. Department of Surgical Biotechnology, Division of Surgery and Interventional Science, UCL, London, UK

Kalliopi Bokea & Maxine Tran

5. Division of Infection and Immunity, Institute of Immunity and Transplantation, UCL, London, UK

Zara Baig

6. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK

Thomas Mitchell

7. Department of Surgery, University of Cambridge, Cambridge, UK

Thomas Mitchell

## Contributions

V.Z. and V.P. provided intellectual input to the project, designed and performed the experiments, performed statistical analysis and generated the figures. I.H.-M. performed the computational analysis of the gene expression datasets. J.J. and I.M.N.W. assisted V.Z. with the generation and characterization of the inducible cell lines. J.L.M. contributed to the acquisition of super-resolution images and assisted V.P. with some confocal-image acquisition. A.F. contributed to the formulation of the project, and performed initial experiments relating to mitochondrial morphology and the observation of cytosolic DNA foci in *FH*-deficient cells. S.R.C. performed super-resolution microscopy analyses of MDVs and assisted V.P. with the quantification of mitochondrial morphology. A.S.H.C., L.T., E.N. and M.Y. performed the mass spectrometry and analysis of the tissue and cell samples. G.C.P. assisted V.P. with the respiration analysis in NDI1 experiments. T.Y. assisted V.Z. with the mouse tissue and cell line processing as well as processing of cell samples for mass spectroscopy. D.B. and S.S. processed the RNA-sequencing raw data for analysis. F.C. performed the electron microscopy imaging acquisition of the mouse kidney tissue. S.N. and M.J. assisted V.P. with immunoblot acquisition. A.S. assisted V.Z. with the in vivo experiments. M.T., K.B. and Z.B. provided HLRCC samples and performed the analyses of inflammatory markers. T.M. performed the deconvolution analyses to

assess the immune contribution in the mouse tissue and human data from bulk gene expression. M.S.-M. generated the cytoFH construct and performed the experiments with the cytoFH-expressing cells. V.Z., V.P., J.P. and C.F. wrote the manuscript with input from all authors. J.P. and C.F. supervised the experiments, conceived and oversaw the study. All authors read and approved the manuscript. Equal second contribution: I.H.-M., J.J., I.M.N.W., J.L.M. and A.F.

### Corresponding authors

Correspondence to Julien Prudent or Christian Frezza.

## Ethics declarations

### Competing interests

C.F. is a scientific adviser for Istesso. The remaining authors declare no competing interests.

## Peer review

### Peer review information

*Nature* thanks Navdeep Chandel, A. Phillip West and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

## Additional information

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

## Extended Data Fig. 1 Characterization of an inducible mouse model of *Fh1* loss.

**a**, Genomic DNA PCR using primers[4] flanking exons 3 and 4 of *Fh1* showing amplification product for the *Fh1* wild-type allele (bottom band; $Fh1^{+/+}$), the floxed (unrecombined) allele (top band; $Fh1^{fl/fl}$) and the recombined allele (middle band; excised LoxP fragments; $Fh1^{-/-}$) from kidney samples at day 5 post-induction. Mice carrying one allele of each $Fh1^{+/+}$ and $Fh1^{fl/fl}$ i.e. $Fh1^{+/fl}$*; Rosa26-Cre-ERT2* were used for comparison with $Fh1^{+/+}$ and $Fh1^{-/-}$ mice and labelled $Fh1^{+/-}$. **b**, Immunoblots of specified proteins in kidney tissue samples at day 5 post-induction. Five kidney samples are shown. **c**, Upon loss of Fh1 activity, fumarate cannot enter the enzymatic reaction that converts it into succinate but, instead, can enter a chemical reaction termed Michael addition[7] whereby it is chemically added to thiol residues of free cysteine to form 2SC, and of proteins which is considered a metabolic marker of FH activity loss. This reaction can be seen as a "buffering tank" that mop-up the excess of fumarate. Therefore, succinated proteins and 2SC levels increase before an intracellular increase in fumarate is detected. Only when succination is "at saturation" with the 2SC intracellular sinks (or "buffering tanks") full, fumarate starts to accumulate. Thus, a modest increase in fumarate can still be accompanied by a strong effect. **d**, Volcano plots showing the differentially expressed genes in $Fh1^{-/-}$ *vs* $Fh1^{+/+}$ kidney at day 5 (left) and day 10 (right). **e**, Deconvolution method on bulk expression data (https://github.com/Danko-Lab/TED and Methods) was applied to determine the cellular composition of mouse kidney tissue at day 5 and day 10 post-induction. Pairwise comparison using Wilcoxon rank sum exact test and Benjamini–Hochberg p-value adjustment were used. **f**, Immunohistological staining of $Fh1^{fl/fl}$ (top) *vs* $Fh1^{-/-}$ (bottom) mouse kidney tissue at day 10 post-induction. Left: 2SC staining, right: CD14 staining. Scale bars: 100 µm.

Source data

## Extended Data Fig. 2 Characterization of inducible *iFh1* epithelial kidney cell lines.

**a**, Schematic diagram of the generation of inducible *iFh1* epithelial kidney cell lines clones 29 ($iFh1^{fl/flCL29}$) and 33 ($iFh1^{fl/flCL33}$). **b**, Mitochondrial membrane potential analysis using MitoTracker Red CMXROS in $iFh1^{fl/flCL29}$ and $iFh1^{fl/flCL33}$ cells treated with either vehicle (ethanol; EtOH) or 4-OHT ($iFh1^{-/-CL29}$ and $iFh1^{-/-CL33}$, respectively). Scale bar: 25 µm. **c**, Mitochondrial respiration measured using Seahorse in $iFh1^{CL29}$ cells. $n = 3$ independent experiments. **d**, qRT-PCR showing expression levels of the transcriptional marker of *Fh1* loss, *Hmox1*, in $iFh1^{CL29}$ cells. $n = 5$ independent experiments. Bar graphs show the fold change expression, for which the expression in control samples was set to 1. Indicated p-values are relative to the corresponding vehicle (EtOH)-treated time point. **e**, Schematic of Pattern Recognition Receptors (PRR) and downstream cascades. **f**, qRT-PCR showing expression levels of *Ifnb1* and ISGs (*Ifi202b*, *Ifitm10*, *Areg* and *Ccl20*) in $iFh1^{CL29}$ cells. $n = 5$ independent experiments. Bar graphs show the fold change expression, for which the expression in control samples was set to 1. Indicated p-values are relative to the corresponding vehicle (EtOH)-treated time point. **g-i**, Classification of mitochondrial morphology (**g**), quantification of mitochondria number (**h**), and area (**i**), in $iFh1^{CL29}$ cells. $n = 3$ independent experiments. Indicated p-values are relative to the corresponding vehicle (EtOH)-treated morphology category. **j**, TEM images (left) and mean mitochondria volume quantification (right) in wild-type ($Fh1^{+/+}$) or chronic floxed *Fh1* (c$Fh1^{fl/fl}$) mouse kidneys epithelial cells treated with Cre-expressing Adenovirus. Data are mean ± s.e.m. **d**,**f**, Students t-test corrected for multiple comparison with the Holm-Sidak method, **g**, two-way ANOVA with Tukey's multiple comparison test, **h**,**i**, one-way ANOVA with Tukey's multiple comparison test. For (**g**) p-values are indicated above each specific phenotype and relative to the corresponding phenotype in vehicle-treated control.

Source data

## Extended Data Fig. 3 Chronic loss of *Fh1* leads to abnormal mitochondrial morphology and mtDNA release in the cytosol.

**a**, Representative N-SIM super-resolution images of mitochondrial morphology (TOM20) in epithelial kidney cell line with chronic *Fh1* deletion ($cFh1^{-/-CL1}$) from $cFh1^{fl/fl}$, compared to $cFh1^{fl/fl}$ cells. Scale bar: 5 µm. **b**, Quantification of mitochondrial diameter in *cFh1* cells from (**a**). $n = 3$ independent experiments. **c**, Quantification of mitochondrial morphology in *cFh1* cells. $n = 3$ independent experiments. **d**, Quantification of mitochondrial number (left) and area (right) in *cFh1* cells. $n = 3$ independent experiments. **e**, Representative confocal images of mitochondrial morphology (TOM20) and DNA foci (DNA) in *cFh1* cells. White arrows indicate cytosolic DNA foci. Scale bar: 10 µm. **f**,**g**, Percentage of *cFh1* cells showing cytosolic DNA (**f**), and number of cytosolic DNA foci per cell (**g**), from **e**. $n = 6$ independent experiments. **h**-**j**, Quantification of mtDNA copy number by ddPCR using either a mtCo3 (**h**), ND1 (**i**) or D-loop (**j**) probe, from isolated cytosolic fractions of *cFh1* cells. $n = 3$ independent experiments. **k**, Immunoblots of specified proteins in *cFh1* cells. **l**, Expression of inflammation-related *Ifi202b* and ISGs in *cFh1* cells measured by qRT-PCR. $n = 3$ independent experiments. Data are mean ± s.e.m. **b**,**d**,**f**-**j**, one-way ANOVA with Tukey's multiple comparison test, **c**,**l**, two-way ANOVA with Tukey's multiple comparison test.

Source data

## Extended Data Fig. 4 MMF treatment phenocopies *Fh1* loss.

**a**, Expression levels of the transcriptional markers of *Fh1* loss *Hmox1* and *Nqo1* as well as *Tfam* expression in monomethyl fumarate (MMF)-treated $cFh1^{fl/fl}$ cells for 8 days compared to vehicle (DMSO)-treated cells, measured by qRT-PCR. $n = 3$ independent experiments. Bar graphs show the fold change expression, for which the expression in control samples was set to 1. **b**, Relative abundance of fumarate levels measured by LC–MS in $iFh1^{fl/flCL29}$ untreated (NT) or treated with vehicle ($iFh1^{fl/flCL29}$) or 4-OHT ($iFh1^{-/-CL29}$) or 400 µM MMF for the indicated period of time. $n = 5$ independent experiments. **c**, Relative abundance of 2SC levels measured by LC–MS in $iFh1^{CL29}$ cells and $iFh1^{fl/flCL29}$ treated with 400 µM MMF for the indicated period of time. $n = 5$ independent experiments. **d**, Representative confocal images of $cFh1^{fl/fl}$ cells treated with vehicle (DMSO) or 400 µM MMF for the indicated period of time. Mitochondria

and succinated proteins were labelled using anti-TOM20 and anti-2SC antibodies, respectively. Blue arrows indicate 2-SC-decorated proteins accumulation in mitochondria. Scale bar: 10 µm. **e**-**g**, Classification of mitochondrial morphology (**e**), quantification of mitochondrial number (**f**), and area (**g**) in *cFh1*$^{fl/fl}$ cells treated with DMSO or 400 µM MMF for 8 days. *n* = 3 independent experiments. Data are mean ± s.e.m. **a**–**c**, one-way ANOVA with Tukey's multiple comparison test, **e**, two-way ANOVA with Tukey's multiple comparison test, **f**,**g**, one-way ANOVA with Tukey's multiple comparison test, for (**e**) p-values are indicated above each specific phenotype and relative to the corresponding phenotype in vehicle (DMSO)-treated control.

Source data

## Extended Data Fig. 5 MMF treatment induces mtDNA release and the translocation of pTBK1 to the Golgi apparatus.

**a**, Percentage of MMF-treated *cFh1*$^{fl/fl}$ cells for 8 days showing cytosolic DNA foci according to mitochondrial morphology phenotype. *n* = 3 independent experiments. **b**,**c**, Quantification of mtDNA copy number by ddPCR using either a ND1 (**b**) or D-loop (**c**) probe, from isolated cytosolic fractions of DMSO- or MMF-treated (8 days) *cFh1*$^{fl/fl}$ cells. *n* = 3 independent experiments. **d**, Representative confocal images of *cFh1*$^{fl/fl}$ treated with 400 µM MMF or DMSO for 8 days. Mitochondria, Golgi apparatus, and pTBK1 were labelled using anti-TOM20, anti-GM130 and anti-pTBK1 antibodies, respectively. Scale bar: 10 µm. **e**, Quantification of *cFh1*$^{fl/fl}$ showing the percentage of cells with pTBK1 recruitment and colocalization with the Golgi apparatus marker, GM130 from **d**. *n* = 3 independent experiments. Data are mean ± s.e.m. **b**,**c**, one-way ANOVA with Tukey's multiple comparison test, **e**, Students paired t-test.

Source data

## Extended Data Fig. 6 Cytosolic re-expression of FH only partially rescues the *Fh1* loss phenotype.

**a**, Representative confocal images of *cFh1*$^{-/-CL1}$ and *cFh1*$^{-/-CL1}$ cells stably expressing pcytoFH-EGFP (*cFh1*$^{-/-CL1}$+*cytoFh1-GFP*). Mitochondria and succinated proteins were labelled using Mitotracker and anti-2SC antibody, respectively. Nucleus was labelled using DAPI. Scale bar: 25 µm. **b**, Representative confocal images of *cFh1*$^{fl/fl}$ and *cFh1*$^{-/-CL1}$ cells stably expressing pcytoFH-EGFP (*cFh1*$^{-/-CL1}$+*cytoFh1-GFP*) or pFH-EGFP (*cFh1*$^{-/-CL1}$+ *pFH-GFP*). Mitochondria and DNA were labelled using anti-TOM20 and anti-DNA antibodies, respectively. White arrows indicate cytosolic DNA foci. Scale bar: 10 µm. **c**, *Fh1* mRNA expression levels in *cFh1* cells measured by qRT-PCR. $n = 3$ independent experiments. **d**, Basal OCR in *cFh1* cells measured using Seahorse. $n = 5$ independent experiments. **e**,**f**, Relative abundance of fumarate (**e**), and 2SC (**f**), in *cFh1* cells measured by LC–MS. $n = 5$ independent experiments. **g**, Number of cytosolic DNA foci in *cFh1* cells from **b**. $n = 3$ independent experiments. **h**, Representative immunoblots of specified proteins in *cFh1* cells. **i**, mRNA expression levels of a panel of ISGs in *cFh1* cells measured by qRT-PCR. $n = 3$ independent experiments. Data are mean ± s.e.m. **c**-**g**,**i**, one-way ANOVA with Tukey's multiple comparison test.

Source data

## Extended Data Fig. 7 Effects of treatment with TCA-cycle intermediates on cytosolic mtDNA and the inflammation profile.

**a**, Representative confocal images of mitochondrial morphology (TOM20) and DNA foci (DNA) in *iFh1*$^{fl/flCL29}$ cells treated with dimethylsuccinate (DMS) or vehicle (DMSO) for 8 days at the indicated concentration. Scale bar: 10 µm. **b**,**c**, Quantification of *iFh1*$^{fl/flCL29}$ cells showing the percentage of cells with cytosolic DNA (**b**), and the number of cytosolic DNA foci per cell (**c**), from **a**. $n = 3$ independent experiments. **d**, Quantification of mtDNA copy number by ddPCR using a mtCo3 probe, from isolated cytosolic fractions of *iFh1*$^{fl/flCL29}$ cells treated with DMSO or DMS. $n = 3$ independent experiments. **e**,**f**, Immunoblots of specified proteins in *iFh1*$^{fl/flCL29}$ cells treated with αKG (**e**), and 2HG (**f**), at indicated concentration. **g**, Expression of the ISGs *Ifnb1*, *Ifi202b*, *Ifitm10*, *Ccl20* and

*Areg* in *iFh1*$^{fl/fl CL29}$ cells treated with DMSO, MMF, αKG or 2HG at indicated concentration, measured by qRT-PCR. *n* = 5 independent experiments. Data are mean ± s.e.m. **b**-**d**,**g**, one-way ANOVA with Tukey's multiple comparison test, **b**-**d**, Bar graphs show the fold change expression, for which the expression in control samples was set to 1. p-values indicated above each bar are relative to the DMSO control.

Source data

## Extended Data Fig. 8 In cellulo and in vivo inhibition of the cGAS–STING pathway.

**a**, Top: schematic showing depletion of mtDNA in *iFh1*$^{fl/fl CL29}$ cells using ethidium bromide (EtBr). Bottom: representative confocal images of untreated *iFh1*$^{fl/fl CL29}$ and mtDNA-depleted *iFh1*$^{fl/fl CL29 \rho 0}$ cells. Mitochondria and DNA were labelled using anti-TOM20 and anti-DNA antibodies, respectively. Scale bar: 10 µm. **b**, mRNA expression of the ISGs *Ifnb1*, *Areg* and *Mx1* in non-induced *iFh1*$^{fl/fl\ CL29}$ and mtDNA-depleted *iFh1*$^{fl/fl\ CL29}$ (*Fh1*$^{\rho 0}$) cell lines, treated with vehicle (DMSO) or 400 µM MMF for the indicated period of time, measured by qRT-PCR. *n* = 3 independent experiments. **c**, Representative confocal images of *cFh1*$^{fl/fl}$ cells treated with 400 µM MMF or vehicle (DMSO) for 8 days. cGAS and nucleus were labelled using an anti-cGAS antibody and DAPI staining, respectively. Scale bar: 10 µm. **d**, Quantification of *cFh1*$^{fl/fl}$ cells showing the percentage of cells with cytosolic cGAS translocation from **c**. *n* = 3 independent experiments. **e**, mRNA expression of a panel of ISGs in *iFh1*$^{fl/fl CL29}$ cells treated with vehicle (DMSO) or MMF and DMSO or cGAS inhibitor RU.521 for the indicated period of time, measured by qRT-PCR. *n* = 5 independent experiments. Bar graphs show the fold change expression, for which the expression in control samples was set to 1. **f**, mRNA expression of the ISGs *Ifi202b* and *Ifitm10* in *iFh1*$^{fl/fl\ CL29}$ cells at day 10 post-induction, treated with vehicle (DMSO) or cGAS inhibitor (RU.521), measured by qRT-PCR. The 4-OHT *vs* vehicle (EtOH) ratios are shown. *n* = 5 independent experiments. Bar graphs show the fold change expression, for which the expression in RU.521 samples was set to 1. **g**, mRNA expression of the ISGs *Ifi202b*, *Ifitm10*, and *Ccl20* in *cFh1*$^{fl/fl}$ *and*

$cFh1^{-/-CL1}$ cells, treated with vehicle (DMSO) or cGAS inhibitor (RU.521), measured by qRT-PCR. $n = 3$ independent experiments. (**h**) mRNA expression of a panel of ISGs in mouse kidney tissue treated with the STING inhibitor H-151, measured by qRT-PCR. $n = 9$ mice per group. Black dots=$Fh1^{+/+}$ + vehicle, red dots = $Fh1^{-/-}$ + vehicle, light blue dots = $Fh1^{-/-}$ + 0.7 mg H-151, dark blue dots = $Fh1^{-/-}$ + 1.4 mg H-151. **i**, Immunoblots of specified proteins in $iFh1^{cl29}$ cells at 15 days post-induction transfected with indicated siRNA. Data are mean ± s.e.m. **b**, Students t-test (FDR two-stage Benjamini, Krieger and Yekutieli) at the corresponding time points, **d**,**f**, multiple Student's t-test, **e**,**g**, two-way ANOVA with Tukey's multiple comparison test, **h**, one-way ANOVA with Dunnett's multiple comparison test.

Source data

## Extended Data Fig. 9 Targeted siRNA screen of key genes involved in mtDNA release and regulation of mitochondrial morphology.

**a**, Number of cytosolic DNA foci per cell in $cFh1^{-/-CL1}$ cells treated with indicated siRNAs. $n = 3$ independent experiments. **b**-**d**, Quantification of mtDNA copy number by ddPCR using either a mtCo3 (**b**), ND1 (**c**) or D-loop (**d**) probe, from isolated cytosolic fractions of $cFh1^{-/-CL1}$ cells. $n = 3$ independent experiments. **e**, Representative confocal images of $cFh1^{+/+}$ and $cFh1^{-/-CL1}$ cells. Mitochondria and cytochrome *c* (Cyt *c*) were labelled using anti-TOM20 and anti-cytochrome *c* antibodies, respectively. Scale bar: 10 µm. **f**, Immunoblots of specified proteins from whole cells (WCE), cytosol and heavy membrane (crude mitochondria) isolated fractions of *cFh1* cells. Data are mean ± s.e.m. **a**-**d**, two-tailed unpaired t-test, p-values are indicated above each specific siRNA and relative to the scramble (scr) siRNA transfected control.

Source data

## Extended Data Fig. 10 Loss of *Snx9* rescues mtDNA release and TBK1–IRF3 activation triggered by *Fh1* loss.

**a**, Representative confocal images of mitochondrial morphology (TOM20) and DNA foci (DNA) in *cFh1*$^{-/-CL19}$ cells transfected with scramble (scr) or *Snx9* siRNA. White arrows indicate cytosolic DNA foci. Scale bars: 10 µm. **b**, Number of cytosolic DNA foci per cell in *cFh1* cells. $n = 3$ independent experiments. **c**, Percentage of cells with cytosolic DNA foci in *c*Fh1 cells. $n = 3$ independent experiments. **d**, Immunoblots of specified proteins of *cFh1* cells. **e**,**f**, Quantification of mitochondrial number (**e**), and area (**f**), in *cFh1* cells. $n = 3$ independent experiments. **g**,**h**, Average number of cytosolic DNA foci per cell (**g**), and number of cytosolic DNA foci per cell (**h**), in *iFh1*$^{fl/flCL29}$ cells transfected with scr or *Snx9* siRNA and treated with either vehicle (EtOH) or 4-OHT (*iFh*$^{-/-CL29}$) for the indicated period of time. $n = 3$ independent experiments. **i**, Immunoblots of specified proteins in *iFh1*$^{CL29}$ cells. **j**-**l**, Quantification of mtDNA copy number by ddPCR using either a mtCo3 (**j**), ND1 (**k**) or D-loop (**l**) probe, from isolated cytosolic fractions of *iFh1*$^{fl/flCL29}$ cells. $n = 3$ independent experiments. Data are mean ± s.e.m. **b**,**c**,**e**-**h**,**j**-**l**, one-way ANOVA with Tukey's multiple comparison test.

Source data

## Extended Data Fig. 11 Characterization of MDVs containing mtDNA.

**a**,**b**, Percentage of cells harbouring TOM20$^{-}$PDH$^{+}$DNA$^{+}$ MDVs (**a**), and number of TOM20$^{-}$PDH$^{+}$ MDVs containing or not mtDNA per cell (**b**), in *cFh1*$^{fl/fl}$ cells treated with vehicle (DMSO) or 400 µM monomethyl fumarate (MMF) for the indicated period of time. $n = 3$ independent experiments. p-values are indicated above each specific time point and relative to the untreated control for each of the one (a) or two (b) parameters. **c**, Number of cytosolic DNA foci per cell in *cFh1*$^{fl/fl}$ cells. $n = 3$ independent experiments. **d**, Representative Airyscan super-resolution images of *cFh1*$^{fl/fl}$ cells treated with 400 µM MMF for 6 days. Mitochondria were labelled using anti-TOM20 and anti-PDH antibodies, TFAM using an anti-TFAM antibody, and DNA using an anti-DNA antibody. White arrows indicate TOM20$^{-}$PDH$^{+}$DNA$^{+}$TFAM$^{+}$ MDVs. Scale bar: 5 µm. **e**, Representative lattice super-resolution SIM image of MMF-

treated *cFh1*$^{fl/fl}$ cells (6 days). Mitochondria were labelled using anti-TOM20 and anti-PDH antibodies, TFAM using an anti-TFAM antibody, and DNA using an anti-DNA antibody. Scale bar: 5 µm; magnification : scale bar: 1 µm. **f**,**g**, Quantification of the average maximal Feret's diameter (**f**), and average area (**g**), of TOM20$^{-}$PDH$^{+}$DNA$^{+}$ MDVs from **e**. $n = 3$ independent experiments. **h**, Representative confocal images of *iFh1*$^{fl/flCL29}$ cells treated with either vehicle (ethanol; EtOH) or 4-OHT (*iFh*$^{-/-CL29}$) for the indicated period of time. Mitochondria were labelled using anti-TOM20 and anti-PDH antibodies, and DNA using an anti-DNA antibody. White arrows indicate TOM20$^{-}$PDH$^{+}$DNA$^{+}$ MDVs. Scale bar: 10 µm. **i**, Quantification of TOM20$^{-}$PDH$^{+}$DNA$^{+}$ MDVs from **h**. $n = 3$ independent experiments. Data are mean ± s.e.m. **a**,**c**,**i**, one-way ANOVA with Tukey's multiple comparison test, **b**, two-way ANOVA with Tukey's multiple comparison test.

Source data

## Extended Data Fig. 12 Loss of *Snx9* prevents MDV formation and inflammation induced by *Fh1* loss.

**a**, Representative confocal images of *iFh1*$^{fl/flCL29\rho0}$ cells treated with 400 µM MMF for 1–8 days. Mitochondria were labelled using anti-TOM20 and anti-PDH antibodies, and DNA using an anti-DNA antibody. Scale bar: 10 µm. **b**, Quantification of the number of TOM20$^{-}$PDH$^{+}$ MDVs from **a**. $n = 3$ independent experiments. **c**, Number of cytosolic DNA foci per cell in *cFh1*$^{fl/fl}$ cells pre-transfected with scramble (scr) or *Snx9* siRNA and treated with vehicle (DMSO) or MMF for 6 days. $n = 3$ independent experiments. **d**, mRNA expression of a panel of ISGs in *cFh1*$^{fl/fl}$ cells pre-transfected with scramble (scr) or *Snx9* siRNA and treated with 400 µM MMF or vehicle (DMSO) for the indicated period of time, measured by qRT-PCR. $n = 3$ independent experiments. Bar graphs show the fold change expression, for which the expression in control samples was set to 1. Data are mean ± s.e.m. **b**-**d**, one-way ANOVA with Tukey's multiple comparison test.

Source data

# Supplementary information

## Supplementary Information

Supplementary Figs. 1–10, including the uncropped Immunoblots (shown as Supplementary Fig. 10) and Supplementary Tables 4 and 5.

## Reporting Summary

## Supplementary Table 1

RNA-sequencing data from the mouse experiment in Fig. 1.

## Supplementary Table 2

Gene-set Analyses using biological processes of the RNA-sequencing experiment in Supplementary Table 1.

## Supplementary Table 3

KEGG pathway analysis of the RNA-sequencing experiment in Supplementary Table 1.

## Supplementary Table 4

List of mouse primers used in the paper in Excel format.

## Supplementary Table 5

List of human primers used in the paper in Excel format.

## Peer Review File

## Source Data

Source data for Supplementary Figs. 1–10.

## Source data

**Source Data Fig. 1**

**Source Data Fig. 2**

**Source Data Fig. 3**

**Source Data Fig. 4**

**Source Data Fig. 5**

**Source Data Extended Data Fig. 1**

**Source Data Extended Data Fig. 2**

**Source Data Extended Data Fig. 3**

**Source Data Extended Data Fig. 4**

**Source Data Extended Data Fig. 5**

**Source Data Extended Data Fig. 6**

**Source Data Extended Data Fig. 7**

**Source Data Extended Data Fig. 8**

**Source Data Extended Data Fig. 9**

**Source Data Extended Data Fig. 10**

[Source Data Extended Data Fig. 11]

[Source Data Extended Data Fig. 12]

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

## About this article

### Cite this article

Zecchini, V., Paupe, V., Herranz-Montoya, I. *et al.* Fumarate induces vesicular release of mtDNA to drive innate immunity. *Nature* **615**, 499–506 (2023). https://doi.org/10.1038/s41586-023-05770-w

- Received: 24 January 2022
- Accepted: 30 January 2023
- Published: 08 March 2023
- Issue Date: 16 March 2023

- DOI: https://doi.org/10.1038/s41586-023-05770-w

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

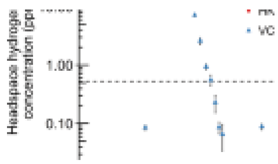
## Mitochondrial molecule controls inflammation

- Taylor A. Poor
- Navdeep S. Chandel

Nature News & Views 08 Mar 2023

---

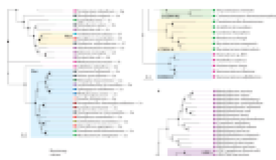
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05770-w




## [The thermoacidophilic methanotroph Methylacidiphilum fumariolicum SolV oxidizes subatmospheric H2 with a high-affinity, membrane-associated [NiFe] hydrogenase]

10 February 2020

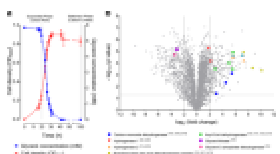
Rob A. Schmitz, Arjan Pol, … Huub J. M. Op den Camp



## [Microbial oxidation of atmospheric trace gases]

12 April 2022

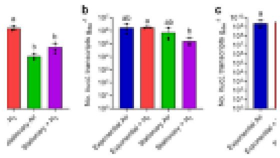
Chris Greening & Rhys Grinter



## [Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival]

29 July 2019

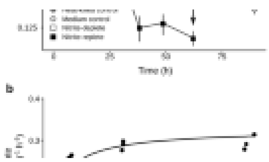
Paul R. F. Cordero, Katherine Bayly, … Chris Greening



## [A widely distributed hydrogenase oxidises atmospheric H2 during bacterial growth]

09 July 2020

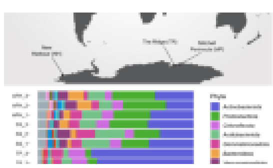
Zahra F. Islam, Caitlin Welsh, … Chris Greening



## [A nitrite-oxidising bacterium constitutively consumes atmospheric hydrogen]

25 June 2022

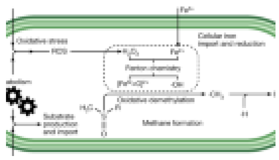
Pok Man Leung, Anne Daebeler, … Chris Greening



## [Atmospheric chemosynthesis is phylogenetically and geographically widespread and contributes significantly to carbon fixation throughout cold deserts]

06 August 2022

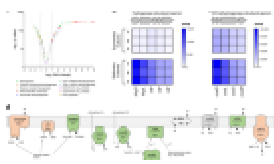
Angelique E. Ray, Julian Zaugg, … Belinda C. Ferrari



### [Methane formation driven by reactive oxygen species across all living organisms]

09 March 2022

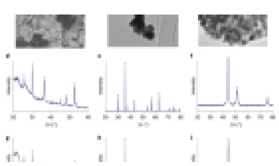
Leonard Ernst, Benedikt Steinfeld, … Frank Keppler



### [Two Chloroflexi classes independently evolved the ability to persist on atmospheric hydrogen and carbon monoxide]

14 March 2019

Zahra F. Islam, Paul R. F. Cordero, … Chris Greening



### [A hydrogen-dependent geochemical analogue of primordial carbon and energy metabolism]

02 March 2020

Martina Preiner, Kensuke Igarashi, … William F. Martin

- Article
- Open Access
- Published: 08 March 2023

# Structural basis for bacterial energy extraction from atmospheric hydrogen


- Rhys Grinter ,
- Ashleigh Kropp,
- Hari Venugopal,
- Moritz Senger,
- Jack Badley,
- Princess R. Cabotaje,
- Ruyu Jia,
- Zehui Duan,
- Ping Huang,
- Sven T. Stripp,
- Christopher K. Barlow,
- Matthew Belousoff,
- Hannah S. Shafaat,
- Gregory M. Cook,
- Ralf B. Schittenhelm,
- Kylie A. Vincent,
- Syma Khalid,
- Gustav Berggren &
- …
- Chris Greening

Show authors



*Nature* **volume 615**, pages 541–547 (2023)



- 72k Accesses
- 1626 Altmetric
- Metrics details

# Abstract

Diverse aerobic bacteria use atmospheric $H_2$ as an energy source for growth and survival[1]. This globally significant process regulates the composition of the atmosphere, enhances soil biodiversity and drives primary production in extreme environments[2,3]. Atmospheric $H_2$ oxidation is attributed to uncharacterized members of the [NiFe] hydrogenase superfamily[4,5]. However, it remains unresolved how these enzymes overcome the extraordinary catalytic challenge of oxidizing picomolar levels of $H_2$ amid ambient levels of the catalytic poison $O_2$ and how the derived electrons are transferred to the respiratory chain[1]. Here we determined the cryo-electron microscopy structure of the *Mycobacterium smegmatis* hydrogenase Huc and investigated its mechanism. Huc is a highly efficient oxygen-insensitive enzyme that couples oxidation of atmospheric $H_2$ to the hydrogenation of the respiratory electron carrier menaquinone. Huc uses narrow hydrophobic gas channels to selectively bind atmospheric $H_2$ at the expense of $O_2$, and 3 [3Fe–4S] clusters modulate the properties of the enzyme so that atmospheric $H_2$ oxidation is energetically feasible. The Huc catalytic subunits form an octameric 833 kDa complex around a membrane-associated stalk, which transports and reduces menaquinone 94 Å from the membrane. These findings provide a mechanistic basis for the biogeochemically and ecologically important process of atmospheric $H_2$ oxidation, uncover a mode of energy coupling dependent on long-range quinone transport, and pave the way for the development of catalysts that oxidize $H_2$ in ambient air.

# Main

The oxidation of atmospheric hydrogen ($H_2$) by soils is a key biogeochemical process that shapes the redox state of the atmosphere[1]. Until recently, this was thought to be an abiotic process, but it is now recognized that diverse aerobic bacteria from at least nine phyla oxidize atmospheric $H_2$ and together account for 75% (around 60 Tg) of the total $H_2$ removed from the atmosphere annually[1,4,6]. Atmospheric $H_2$ oxidation provides bacteria with a supplemental energy source in nutrient-limited soil environments,

enabling them to either grow mixotrophically[7,8,9,10,11] or persist on air alone in a dormant but viable state for long periods[2,4,6,12,13,14,15]. For example, *Mycobacterium* cells and *Streptomyces* spores survive starvation by transferring electrons through an aerobic respiratory chain from atmospheric $H_2$ to $O_2$ (refs. [7,14,16,17]). The ability to oxidize atmospheric $H_2$ is widespread in bacteria from diverse environments[2], and some ecosystems—such as hyper-arid polar soils—appear to be driven primarily by atmospheric energy sources[1,2,3,15].

There are no known chemical catalysts that can oxidize atmospheric $H_2$; this would require the selective oxidation of low concentrations of substrate (530 parts per billion by volume (ppbv)) present in an atmosphere containing a high concentration (21%) of the catalytic poison $O_2$ (refs. [18,19]). Group 1 [NiFe] hydrogenases constitute a family of membrane-bound $H_2$-oxidizing metalloenzymes that support aerobic and anaerobic growth of bacteria in $H_2$-rich environments; however, these enzymes are generally incapable of atmospheric $H_2$ oxidation, given they have a low affinity for $H_2$ (Michaelis constant ($K_m$) >500 nM) and are either reversibly or irreversibly inhibited by $O_2$ (refs. [20,21,22,23,24]). Recently, several high-affinity lineages of group 1 and 2 [NiFe] hydrogenases have been identified that input atmospheric $H_2$-derived electrons into the aerobic respiratory chain[3,4,5,6]. Whole-cell studies suggest that these enzymes have significantly higher apparent affinities for $H_2$ ($K_m$ values 30 to 200 nM) and appear insensitive to inhibition by $O_2$ (refs. [4,6,10,25,26]). However, given these hydrogenases have yet to be isolated, it remains unknown how they have evolved to selectively oxidize $H_2$, tolerate exposure to $O_2$ and interact with the electron transport chain. Notably, it is debated whether the hydrogenases responsible for the oxidation of atmospheric $H_2$ have an inherently high affinity or whether their affinity is modulated by their interactions with the respiratory chain[1,23].

To address these knowledge gaps, we investigated the structural and mechanistic basis of atmospheric $H_2$ oxidation in the aerobic bacterium *M. smegmatis*. This bacterium possesses 2 phylogenetically distinct hydrogenases—Huc (group 2a) and Hhy (group 1h)—that both oxidize $H_2$ to

sub-atmospheric levels[4,17,26]. We isolated Huc directly from *M. smegmatis* and determined the structural and biochemical basis for its oxidation of atmospheric $H_2$.

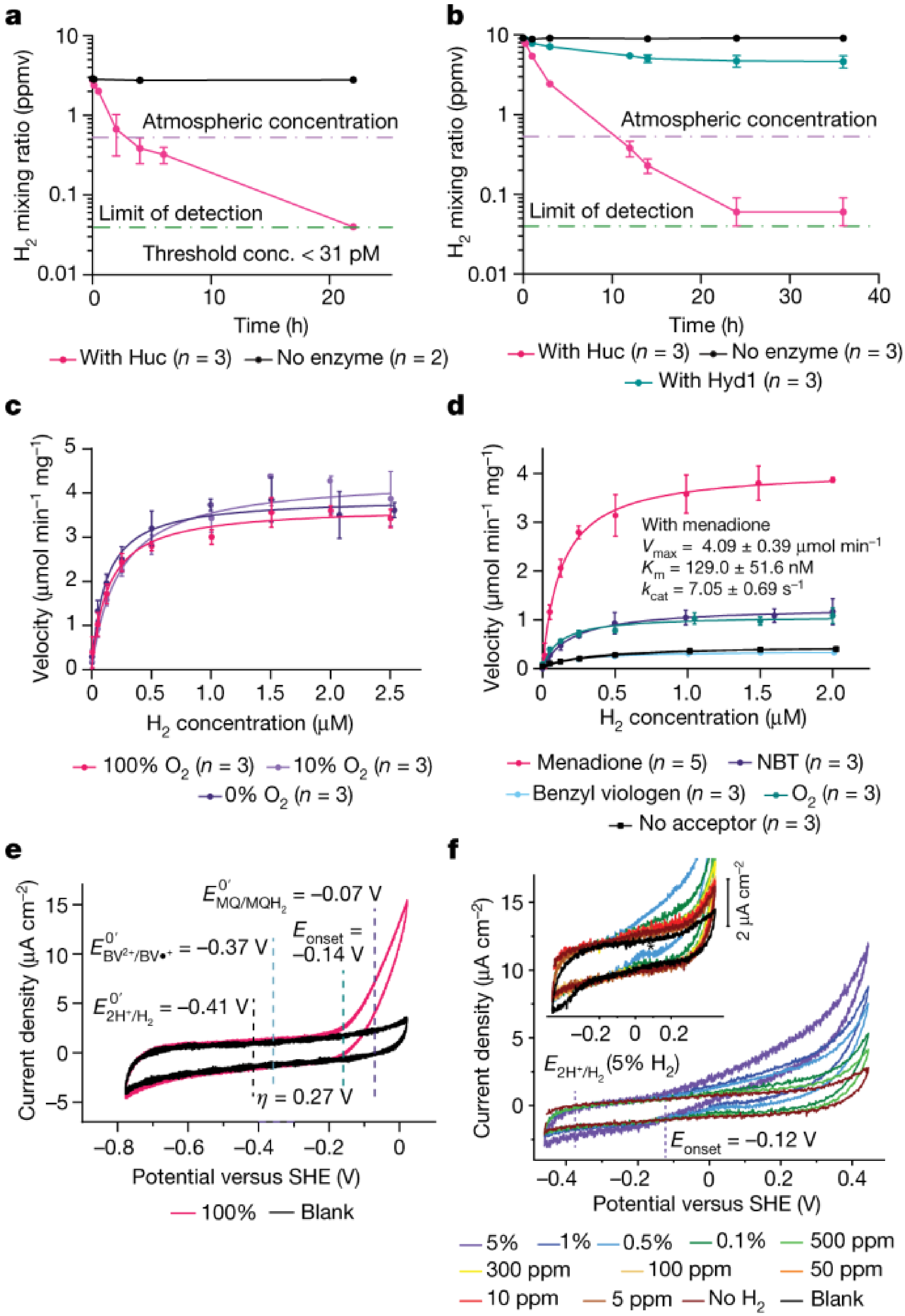
## Huc reduces menaquinone analogues in air

We isolated Huc from *M. smegmatis* using a chromosomally encoded Strep-tag II on its HucS subunit. Analysis by SDS–PAGE showed that Huc consists of three protein subunits corresponding to HucL (approximately 58 kDa), HucS–2×Strep (approximately 39 kDa) and a third unknown subunit (approximately 18 kDa) (Extended Data Fig. 1a). Mass spectrometry identified this third subunit as the uncharacterized protein MSMEG_2261, which is encoded by an open reading frame in the *huc* operon directly upstream of *hucS*[17] (Extended Data Fig. 1b). The co-purification of MSMEG_2261 with the large and small hydrogenase subunits suggests that it is an additional component of the Huc complex, which we designate HucM. As previously observed for Huc in *M. smegmatis* cell lysates, the purified Huc complex migrates on a native gel at a molecular mass of around 800–900 kDa (Extended Data Fig. 1c)[26]. Huc is red–brown in colour, consistent with the presence of multiple metal clusters, and is highly stable at room temperature with a melting temperature of 78.3 °C (Extended Data Fig. 1d). To investigate the role of HucM in the Huc complex, we deleted the *hucM* gene from *M. smegmatis*. Activity staining of cell lysates shows that although Huc is active in this mutant, the high molecular weight species was not present, suggesting that HucM is critical for the formation of the Huc oligomer (Extended Data Fig. 1e).

Although Huc can oxidize atmospheric $H_2$ in whole cells, it was unclear whether it has an inherently high substrate affinity or if this property results from coupling to the respiratory electron transport chain[4]. To resolve this question, we tested the ability of purified Huc to oxidize 3 to 100 parts per million (ppm) $H_2$ in an ambient air headspace. Huc rapidly consumed $H_2$ to below the limit of detection of the gas chromatograph (about 40 ppbv) using nitro blue tetrazolium (NBT) or the quinone analogue menadione as the electron acceptor (Fig. 1a,b and Extended Data Fig. 1f). By contrast, the model $O_2$-tolerant group 1d [NiFe] hydrogenase Hyd1 from *Escherichia*

*coli*[27,28] operated more slowly under these conditions and was unable to oxidize $H_2$ to below 4 parts per million by volume (ppmv) (Fig. 1b and Extended Data Fig. 1f). Kinetic analysis of purified Huc indicates that it is adapted to oxidize atmospheric $H_2$, with a high affinity ($K_m$ = 129 nM) and low $H_2$ threshold (<31 pM) but slow turnover (catalytic constant ($k_{cat}$) = 7.05 $s^{-1}$) (Fig. 1a,d and Supplementary Table 1). The $k_{cat}/K_m$ of Huc is $5.4 \times 10^7$ $M^{-1}$ $s^{-1}$, indicating the enzyme is highly efficient at low $H_2$ concentrations. The kinetics of purified Huc are similar to that of *M. smegmatis* mutants harbouring Huc as the sole hydrogenase ($K_m$ = 184 nM, threshold = 133 pM), suggesting that the hydrogenase itself is the primary determinant of whole-cell $H_2$ affinity[4]. The ability of Huc to oxidize $H_2$ at low concentrations in ambient air also indicates a high tolerance to inhibition by $O_2$. To assess the extent of Huc $O_2$ tolerance, we amperometrically measured its rate of $H_2$ oxidation in buffer containing $O_2$ at 0%, 10% (around 0.12 mM) and 100% (around 1.2 mM) saturation. We found no significant differences in the rate or affinity of Huc-mediated $H_2$ oxidation with varying $O_2$ concentrations, indicating that Huc is an $O_2$-insensitive enzyme (Fig. 1c). In the absence of an alternative electron acceptor, $O_2$ stimulates $H_2$ oxidation by Huc (Fig. 1d), indicating the enzyme can transfer electrons directly to $O_2$.

**Fig. 1: Huc is a high-affinity $O_2$-insensitive hydrogenase.**



a
H2 mixing ratio (ppmv)
Atmospheric concentration
Limit of detection
Threshold conc. < 31 pM
Time (h)
With Huc (n = 3)
No enzyme (n = 2)
b
H2 mixing ratio (ppmv)
Atmospheric concentration
Limit of detection
Time (h)
With Huc (n = 3)
No enzyme (n = 3)
With Hyd1 (n = 3)
c
Velocity (μmol min−1 mg−1)
H2 concentration (μM)
100% O2 (n = 3)
10% O2 (n = 3)
0% O2 (n = 3)
d
Velocity (μmol min−1 mg−1)
With menadione
Vmax = 4.09 ± 0.39 μmol min−1
Km = 129.0 ± 51.6 nM
kcat = 7.05 ± 0.69 s−1
H2 concentration (μM)
Menadione (n = 5)
NBT (n = 3)
Benzyl viologen (n = 3)
O2 (n = 3)
No acceptor (n = 3)
e
Current density (μA cm−2)
E0′MQ/MQH2 = −0.07 V
E0′BV2+/BV•+ = −0.37 V
Eonset = −0.14 V
E0′2H+/H2 = −0.41 V
η = 0.27 V
Potential versus SHE (V)
100%
Blank
f
Current density (μA cm−2)
2 μA cm−2
E2H+/H2 (5% H2)
Eonset = −0.12 V
Potential versus SHE (V)
5%
1%
0.5%
0.1%
500 ppm
300 ppm
100 ppm
50 ppm
10 ppm
5 ppm
No H2
Blank

**a**, Gas chromatography analysis of the $H_2$ concentration of the headspace of sealed vials containing Huc or no enzyme, with 200 µM menadione as the electron acceptor. Huc can oxidize $H_2$ below atmospheric concentrations (violet line) to the limit of detection (green line). conc., concentration. **b**, Gas chromatography analysis of vials containing Huc, *E. coli* Hyd1, or no enzyme, with 200 µM NBT as the electron acceptor. Huc—but not Hyd1—can oxidize $H_2$ at sub-atmospheric levels. **c**, Michaelis–Menten kinetics of $H_2$ oxidation by Huc in buffer with different percentages of $O_2$ saturation, with 200 µM menadione (the 0% $O_2$ comprises degassed buffer with traces of $O_2$). **d**, Michaelis–Menten kinetics of Huc $H_2$ consumption in buffer with different electron acceptors, degassed but with traces of $O_2$. In **a**–**d**, data are mean ± s.d. $V_{max}$, maximum velocity. **e**, Cyclic voltammogram of an immobilized Huc protein film in a 100% $H_2$ atmosphere. $E_{onset}$ is the Huc onset potential determined from the average of four experiments (a single representative curve is shown for clarity). $\eta$ is the estimated Huc overpotential compared to the $2H^+/H_2$ redox couple potential ($({E}_{{2{\rm{H}}}^{+}/{{\rm{H}}}_{2}}^{{0}^{{\prime} }})$). The midpoint potentials of benzyl viologen ($({E}_{{{\rm{BV}}}^{2+}/{\rm{BV}}{\cdot }^{+}}^{{0}^{{\prime} }})$) and menaquinone ($({E}_{{\rm{MQ}}/{{\rm{MQH}}}_{2}}^{{0}^{{\prime} }})$) are indicated. SHE, standard hydrogen electrode. **f**, Cyclic voltammogram of an immobilized Huc protein film with a gas mixture flushing through the headspace comprising argon, $N_2$ and the indicated level of $H_2$. The dashed vertical lines indicate the potential of the $2H^+/H_2$ couple and Huc onset potential, calculated for 5% $H_2$ at ambient pressure. Scans were recorded in the order of low to high $H_2$. The inset plot shows a zoomed view of the voltage response of Huc at low $H_2$ partial pressures.

Source data

In *M. smegmatis* cells, electrons from $H_2$ oxidation by Huc enter the respiratory electron transport chain via the reduction of menaquinone, the major respiratory quinone of mycobacteria[26], but it is not known whether this reduction occurs directly or via an unknown membrane protein

intermediate. To test whether menaquinone is the immediate electron acceptor of Huc, we determined its enzyme kinetics with either menadione (representing the redox head group of menaquinone) or the redox-active dyes NBT or benzyl viologen as electron acceptors. Consistent with menaquinone acting as its physiological electron acceptor, Huc displayed a much higher maximum velocity with menadione ($E^{0'} = -74$ mV) as its electron acceptor than NBT ($E^{0'} = -80$ mV), despite their similar midpoint potentials[29,30] (Fig. 1d and Supplementary Table 1). The low-potential electron acceptor benzyl viologen ($E^{0'} = -374$ mV) did not induce $H_2$ oxidation compared with the buffer control, indicating it does not accept electrons from Huc[31] (Fig. 1d). This is consistent with the role of Huc in oxidizing $H_2$ at low concentrations, as the higher redox potential of the $2H^+/H_2$ redox couple under ambient conditions ($E^{0'} = -136$ mV) would preclude the reduction of low-potential substrates. This suggests Huc is tuned to higher redox potentials.

To probe the potential dependence of Huc activity, we performed cyclic voltammetry on Huc films over a wide range of $H_2$ partial pressures (Fig. 1e,f). At lower partial pressures of $H_2$, the electrocatalytic current is observable above the background capacitive current down to 10 ppm $H_2$ (Fig. 1f). Huc maintained an overpotential of 255 ± 10 mV over the range of $H_2$ partial pressures tested (Extended Data Fig. 2a). For example, at 1 μM $H_2$, at which menadione reduction is close to substrate saturation by $H_2$, the onset potential ($E_{onset}$) measured for Huc is −80 mV ± 10 mV, just negative of the menadione potential ($E^{0'} = -74$ mV), which itself would be sensitive to the ratio of oxidized to reduced species. This is consistent with the role of Huc in oxidizing $H_2$ at low concentrations and would provide selectivity for the reduction of menaquinone over lower potential electron acceptors. Huc shows the greatest overpotential for a [NiFe] hydrogenase recorded to date[27,32,33,34] (see Extended Data Fig. 2b,c for comparison to *E. coli* Hyd1 and Hyd2). No significant current attributable to $H^+$ reduction by Huc was observed, and this catalytic irreversibility is consistent with the enzyme's high overpotential for $H_2$ oxidation (Fig. 1e). At higher levels of $H_2$, Huc does not show evidence of oxidative inactivation or reductive re-activation that are characteristic of other [NiFe] hydrogenases (Extended Data Fig.

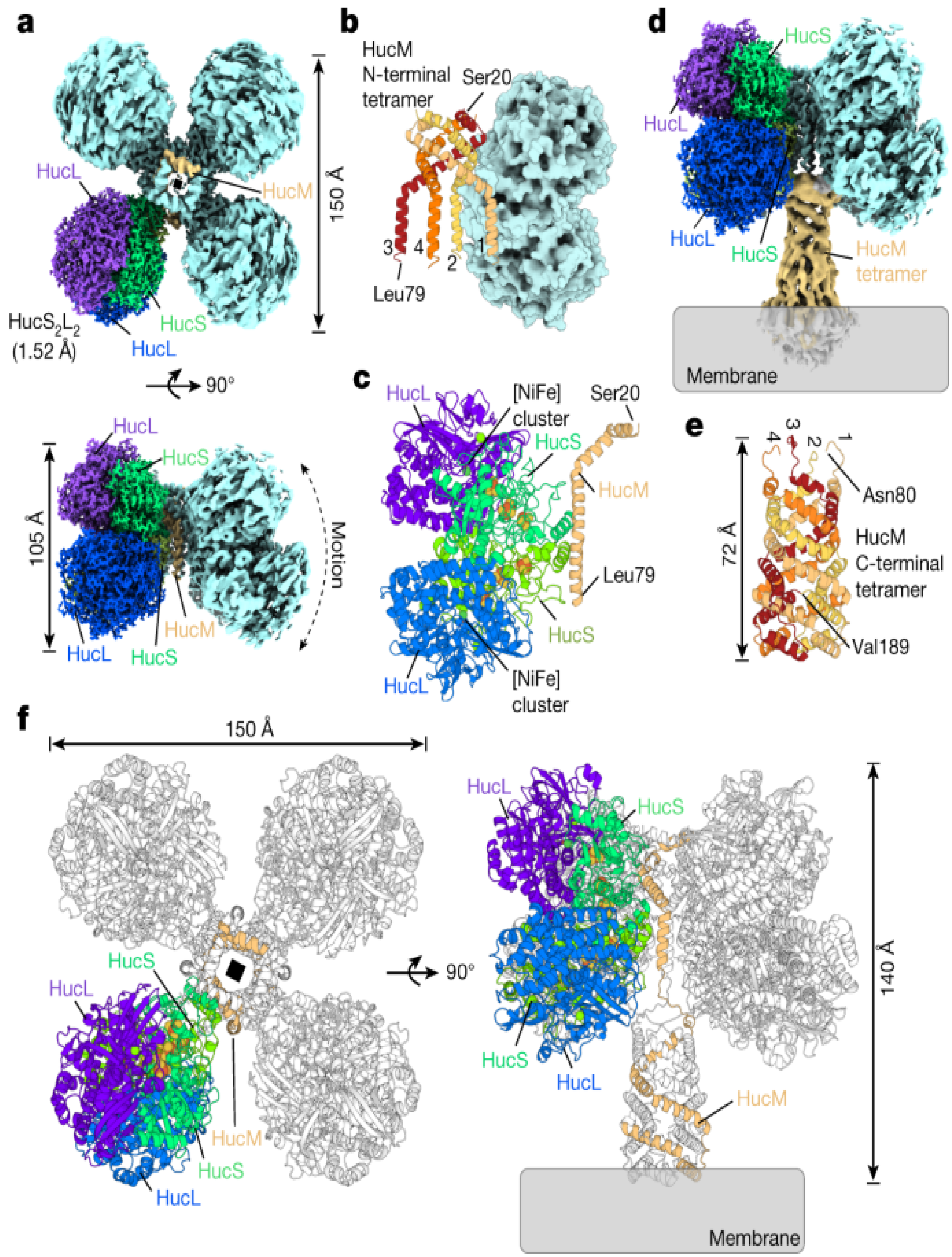
2b,c), although there is evidence of a slight re-activation peak at 0.5% $H_2$ (indicated by the asterisk in Fig. 1e, inset).

## Huc oligomerizes around a central stalk

To determine the structure of Huc, we performed cryo-electron microscopy (cryo-EM) imaging and single-particle reconstruction of the purified 833 kDa complex. Cryo-EM micrographs and resulting class averages revealed a molecule shaped like a four-leaf clover, which is sometimes associated with membrane vesicles via a stalk-like protrusion (Extended Data Fig. 3a,b). In the Huc-overexpressing strain used for purification (see Methods for details), Huc activity is associated mainly with the soluble fraction, yet previous reports indicate that Huc predominantly associates with the cellular membrane[4,26], which is consistent with the presence of membrane-associated Huc in our sample. Three-dimensional reconstruction of the Huc complex yielded maps at a 2.19 Å overall resolution (Extended Data Fig. 3c–e and Extended Data Table 1), showing that each of the four lobes of Huc are composed of two Huc protomers, each consisting of a HucS and HucL subunit (Fig. 2a). The four Huc lobes are bound to a scaffold formed by four molecules of HucM, an elongated α-helical protein that intertwines to form a cage-like structure (Fig. 2b and Extended Data Fig. 4a,b). Because of the flexibility exhibited by this complex (Extended Data Fig. 3f and Supplementary Video 1), we performed 3D reconstruction of the individual Huc lobe from two datasets, obtaining resolutions of 1.52 Å and 1.67 Å (Fig. 2c, Extended Data Fig. 4c,d, Extended Data Table 1 and Supplementary Figs. 2 and 3). The HucSL heterodimer is similar to that of other [NiFe] hydrogenases, with a root mean squared deviation (rmsd) of 3.7 Å and 2.8 Å to hydrogenases from *Desulfovibrio vulgaris* Miyazaki F (Protein Data Bank (PDB) ID = 4U9H) and *Cupriavidus necator* H16 (PDB ID 5AA5), respectively[23,24]. Consistently, the architecture of the [NiFe] active site is the same as previously reported[35] (Fig. 3a). Density maps and the fourth coordinating residue of the 3 iron–sulfur clusters of HucS indicate they all have a [3Fe–4S] configuration (Fig. 3b, Extended Data Fig. 5a and Supplementary Videos 2–4). This is supported by electron paramagnetic resonance (EPR) analysis of Huc, which yields a signal consistent with the presence of multiple spin-coupled [3Fe–4S] clusters and demonstrates their

redox activity (Fig. 3c, Extended Data Fig. 5b and Supplementary Note 1). This is in contrast to other characterized hydrogenases, including $O_2$-tolerant variants, that possess one or more [4Fe–4S] clusters[21,35]. [3Fe–4S] clusters generally undergo their +/0 redox transition at approximately +250 mV higher potential than $[4Fe–4S]^{2+/+}$ clusters[36]. As the potential of the electron transfer chain has previously been shown to dictate catalytic bias in other hydrogenases, the presence of exclusively [3Fe–4S] clusters in Huc probably has a major role in tuning the overpotential of the enzyme[37].

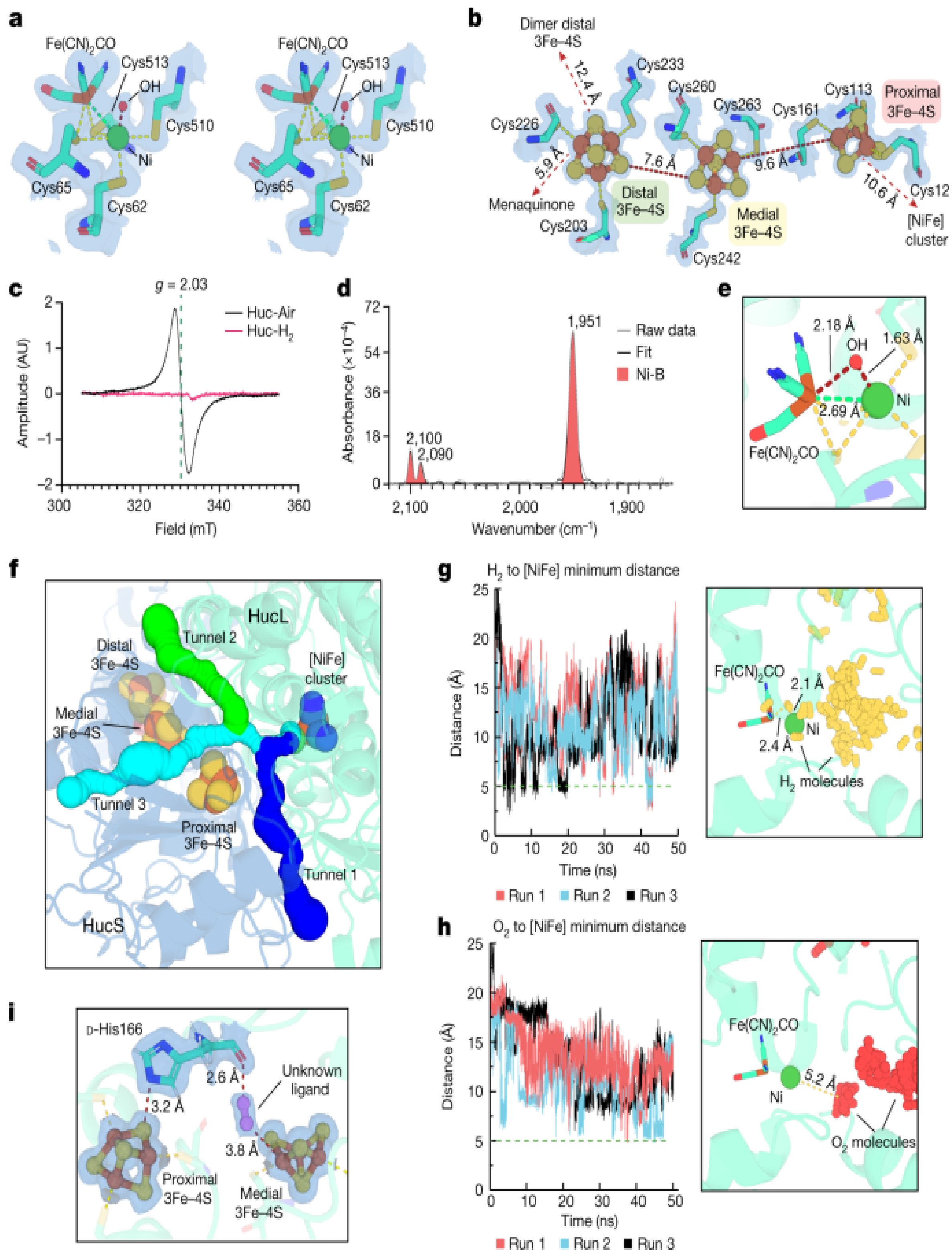
**Fig. 2: Huc forms an 833-kDa oligomer comprising HucS, HucL and HucM.**




**a**, Cryo-EM density maps of the Huc oligomer showing its four lobes that each contain two HucSL dimers, with four centrally located HucM subunits

acting as a scaffold for the oligomer. One lobe of Huc shows the 1.52 Å high-resolution map, with HucL and HucS subunits coloured individually as indicated. One HucM molecule is coloured orange. **b**, A cartoon representation of the central tetramer formed by 4 HucM subunits (residues 20 to 79), with a single $HucS_2L_2$ lobe shown as a surface model for context. **c**, A cartoon representation of the asymmetric unit of the Huc oligomer. **d**, A composite cryo-EM density map showing the Huc 'body' region as in **a**, and the 'stalk' region that is peripherally associated with the cell membrane. **e**, A cartoon representation of the AlphaFold2 model of the C-terminal region of HucM (residues 80 to 189), showing that it forms a tetrameric coiled-coil tube. **f**, A cartoon representation showing the full Huc complex reconstructed from the cryo-EM structure and AlphaFold model of the HucM C-terminal fitted to the cryo-EM density map of the stalk region.

**Fig. 3: Huc adopts a non-inhibitory Ni-B state under ambient conditions and restricts $O_2$ diffusion to the catalytic cluster.**




**a**, A stereo view of the Huc [NiFe] cluster under ambient air. **b**, The structure of the three HucS [3Fe–4S] clusters. **c**, High-temperature (30K)

EPR spectra of Huc in the as-isolated oxidized state (Huc-Air) and the $H_2$-reduced state (Huc-$H_2$). The isotropic signal at $g = 2.03$ ($g$ is the Landé factor) is associated with the oxidized $[3Fe–4S]^+$ clusters, which disappears in Huc-$H_2$ owing to reduction of the clusters. No signal attributable to reduced $[4Fe–4S]^+$ clusters is discernible in the $H_2$-treated form of Huc. **d**, FTIR spectra of oxidized Huc isolated in ambient air, showing that the [NiFe] active site of the enzyme adopts an oxidized state consistent with an Ni-B state. **e**, The Huc [NiFe] active site, showing the relative distances between the Ni-B state hydroxyl ligand and the Ni and Fe ions of the cluster. **f**, The HucSL dimer, showing the location and width of the hydrophobic gas channels that provide substrate access to the [NiFe] active site. **g**, Left, plot of distance versus time, showing the closest distance between $H_2$ and the Huc [NiFe] active site during molecular dynamics simulations of HucSL in the presence of $H_2$. Right, the location of $H_2$ molecules in a representative subset of simulation frames when $H_2$ molecules are in the closest proximity to the active site. **h**, Similar to **g**, but shows location of $O_2$ relative to the Huc [NiFe] site in simulations of HucSL performed in the presence of $O_2$. **i**, The proximity of HucL d-His166 to the proximal and medial [3Fe–4S] clusters and the location of the unknown ligand. Cryo-EM density potential maps in **a**,**b**,**i** are shown as a transparent blue surface at $5\sigma$.

Source data

Out of the 188 amino acids of HucM, only residues 20–79 could be modelled into the cryo-EM density maps, with diffuse density corresponding to the remaining C-terminal region of HucM forming the stalk-like protrusion. To better resolve this region of the structure, we masked it and performed signal subtraction on the remainder of the Huc structure, followed by focused refinement and 3D classification (Extended Data Table 1 and Supplementary Fig. 4). This processing showed that the C-terminal region of HucM forms a tube connecting the main body of the Huc oligomer to the membrane (Fig. 2d and Extended Data Fig. 4e,g). AlphaFold modelling shows this tube consists of a symmetrical coiled-coil of α-helices composed of the four HucM molecules that encloses a chamber at the centre of Huc (Fig. 2e and Extended Data Fig. 4f,g). These α-helices are amphipathic, with the hydrophobic face pointing inwards, lining the tube entirely with

hydrophobic residues (Extended Data Fig. 4h). Interaction of the tube with the phospholipid head groups of the membrane is facilitated by an outward-facing helix at the tube terminus containing arginine, lysine and tryptophan residues (Extended Data Fig. 4i). By docking the model of the HucM tetramer into the map for the Huc stalk, we resolved the structure of the complete Huc complex (Fig. 2f and Supplementary Video 5).

# Unique features enable ambient $H_2$ catalysis

Huc is inherently capable of $H_2$ oxidation at sub-atmospheric concentrations and is largely insensitive to inhibition by $O_2$ (Fig. 1c). This is remarkable for a hydrogenase and highly desirable for the development of biocatalysts that oxidize $H_2$ under aerobic conditions[38,39]. Although Huc may use reverse electron flow to reduce $O_2$ bound at the active site to $H_2O$, it cannot behave as previously described $O_2$-tolerant hydrogenases (for example, *E. coli* Hyd-1), as it lacks a proximal [4Fe–3S] cluster to enable the rapid transfer of four electrons[21,22,40] (Supplementary Note 2). As in several other hydrogenases[23,41], the distal Fe–S clusters of the $HucS_2L_2$ dimer are within electron transfer distance (Figs. 2c and 3b), which could provide a source of electrons from $H_2$ oxidation by the neighbouring dimer subunit to reduce bound $O_2$ (refs. [23,41]). However, as Huc operates under atmospheric conditions, where $O_2$ is four million times more abundant than $H_2$ (refs. [18,19]), it is unlikely that $H_2$ oxidation alone would keep the Huc dimers sufficiently reduced to sustainably displace bound $O_2$, meaning the enzyme must resist $O_2$ inhibition by additional mechanisms. Concordantly, analysis of the hydrophobic gas channels of Huc that provide $H_2$ access to the active site indicates that they are markedly narrower than those of structurally characterized $O_2$-sensitive and $O_2$-tolerant [NiFe] hydrogenases (Fig. 3f, Extended Data Fig. 6a and Supplementary Table 2). Instead, they are of similar width to the $O_2$-insensitive but low-affinity group 1h [NiFe] hydrogenase from *C. necator*[23] (Supplementary Table 2). To determine whether these channels have a role in sterically excluding $O_2$ from the Huc active site, we performed all-atom molecular dynamics simulations on the

HucSL dimer in the presence of excess $H_2$ and $O_2$ dissolved in water. Of note, during these simulations $H_2$ enters the Huc active site, whereas $O_2$ is sterically excluded by a series of bottlenecks between the active site and enzyme surface and does not get closer than 5 Å to the catalytic cluster (Fig. 3g,h and Extended Data Fig. 6). The simulations suggest that the critical point for selection against $O_2$ is a bottleneck after the convergence of the three gas tunnels immediately preceding the active site entrance. In support of this observation, $O_2$ reached the active site in simulations of Huc variants with mutations that relieve this bottleneck (Extended Data Fig. 7 and Supplementary Note 3). This suggests that, as previously proposed for [NiFe] hydrogenases[42], the Huc gas channels have a role in protecting the [NiFe]-cluster from inactivation by $O_2$.

Although the narrow hydrophobic gas channels greatly kinetically favour $H_2$ diffusion over $O_2$, Fourier transform infrared (FTIR) spectroscopy analysis and the cryo-EM structure indicate that the active site of Huc isolated in ambient air is in an oxidized state with a bridging hydroxide ligand (Fig. 3a,d and Supplementary Note 2). The FTIR vibrational frequencies are consistent with an electron-deficient Ni-B species (Extended Data Fig. 8a,b). This is further supported by EPR spectra collected of air-exposed Huc, which shows a rhombic signal that closely resembles the canonical Ni-B species[32] (Extended Data Fig. 5c,d and Supplementary Note 1). Oxidation of the enzyme by $O_2$ resulting in this Ni-B state probably occurs via an indirect pathway, in line with previous studies showing the $O_2$-independent formation of Ni-B states[43,44]. The FTIR spectra support this hypothesis: when reduced Huc is exposed to an atmosphere containing 20% $O_2$, the Ni-SI state is rapidly populated at the expense of more reduced Ni-R and Ni-C states, followed by the slow appearance of the hydroxide-bound Ni-B (Extended Data Fig. 8e,f and Supplementary Note 2).

The high-affinity kinetics of Huc differ radically from more classical $O_2$-sensitive hydrogenases, which tend to be faster low-affinity enzymes[4,45] (Fig. 1d). A comparison of the environment surrounding the [NiFe] active site of Huc and $O_2$-sensitive hydrogenases reveals no differences that could readily explain the divergent properties of these enzymes (Extended Data

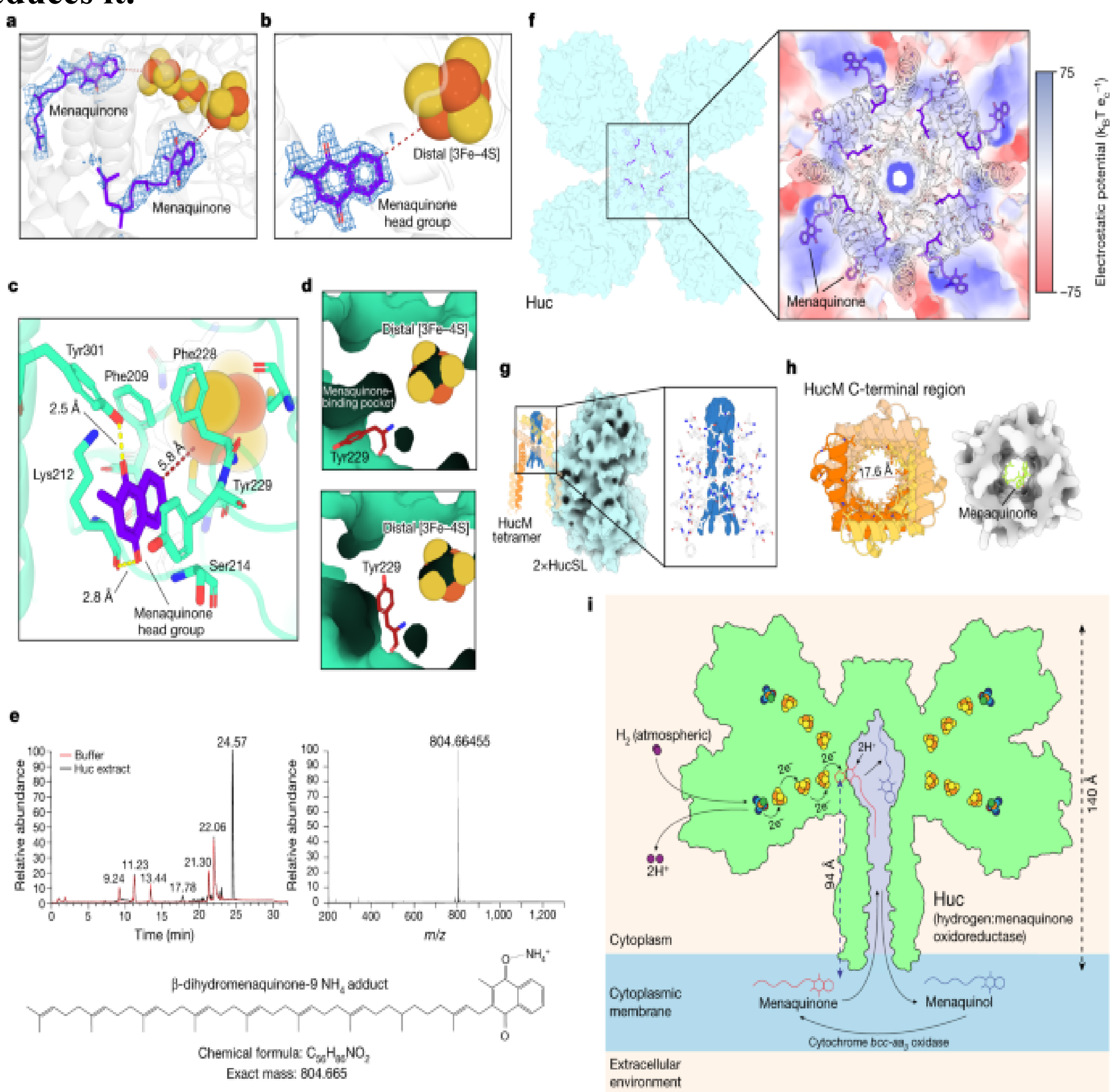
Fig. 9a–c). This indicates that the overall activity of the [NiFe] active site instead results from changes in the properties of the iron–sulfur clusters of the small subunit or other regions of the enzyme. Given their distinct structure, the modified gas channels identified in Huc may have a role in selectively capturing $H_2$ molecules and delivering them to the active site, decreasing the rate but increasing the enzyme's affinity[42]. In addition, the increased redox potential of Huc ($E_{onset} \approx -80$ to $-160$ mV) (Fig. 1e,f) means that atmospheric $H_2$ oxidation, with a redox potential ($E_{atm}$) of −0.134 mV at pH 7, is thermodynamically favourable in comparison to low-affinity variants operating at minimal overpotential[46] ($E_{onset} \approx -360$ mV). The redox properties of the Huc proximal and medial clusters may be further modified by an unusual d-isomer of histidine, indicated by the cryo-EM maps, at position 166 of HucL (Extended Data Fig. 9d and Supplementary Videos 6 and 7). An l-isomeric histidine is conserved at this position in other hydrogenases, with its imidazole head group interacting with the proximal [Fe–S] cluster (Extended Data Fig. 9e). In Huc, His166 also interacts with the proximal cluster (Fig. 3i), but the d-isomeric configuration shifts the position of the backbone carbonyl of His166 so that it interacts with a highly coordinated ligand in proximity to the medial cluster (Fig. 3i and Extended Data Fig. 9f). We were unable to identify this ligand, although its elongated density suggests that it is not water, and it may have a role in tuning the redox potential of the medial [3Fe–4S] cluster.

# Huc reduces menaquinone internally

Analysis of the Huc EM maps reveals additional density near the distal [3Fe–4S] cluster of each Huc protomer, which could be unambiguously modelled as menaquinone (Fig. 4a,b and Supplementary Video 8). Consistent with this, mass spectrometry confirmed that β-dihydromenaquinone-9 associates with purified Huc with an occupancy of 96.2% (Fig. 4e and Supplementary Fig. 5). The head group of menaquinone is stabilized by hydrogen bonding between its oxygen groups and the terminal hydroxyl of Tyr301 and the backbone carbonyl of Lys212, as well as π–π interactions with Tyr229 of HucS (Fig. 4c). In our 1.67 Å reconstruction, Tyr229 adopts a second conformation that clashes with density corresponding to menaquinone, which shields the proximal [3Fe–4S]

cluster from the solvent and may mitigate $O_2$ reduction in the absence of menaquinone, which would lead to the formation of reactive oxygen species (Fig. 4d and Supplementary Video 9). These data indicate that Huc directly binds and reduces menaquinone, which is consistent with menadione acting as the preferred electron acceptor for Huc (Fig. 1c). This contrasts with observations made in group 1 [NiFe] hydrogenases, where small subunits are electron relays without catalytic roles[20,21,22,23,24].

**Fig. 4: Huc extracts menaquinone from the membrane and directly reduces it.**



**a**, Density corresponding to menaquinone bound to Huc from low-resolution cryo-EM reconstruction of the Huc oligomer. **b**, Density corresponding to bound menaquinone in high-resolution cryo-EM maps of the Huc dimer. Density for the menaquinone tail is absent in these maps owing to masking and symmetry averaging. **c**, Coordination of the menaquinone head group by the HucS subunit within electron transfer distance of the distal [3Fe–4S] cluster. **d**, The two conformations observed for Tyr229 of HucS in the high-resolution Huc cryo-EM density maps. Tyr229 adopts an open conformation in the presence of menaquinone (top) and a closed conformation (bottom) that is mutually exclusive with menaquinone binding. **e**, High performance liquid chromatography–mass spectrometry (LC–MS) analysis of the Folch extract from purified Huc. The base peak chromatogram (left) shows a substantial peak at 24.6 min corresponding to an ion at $m/z = 804.66455$ (right), which is consistent with the ammonium adduct of β-dihydromenaquinone-9, $m/z = 804.66531$ (bottom). **f**, A hybrid stick and electrostatic surface representation of the central cavity Huc, showing that the surface of the chamber is lined with hydrophobic residues. **g**, A view of the HucM tetramer, showing the presence of density for lipids that occlude the top of the Huc internal hydrophobic chamber. **h**, A top-down view of the HucM C-terminal tube, showing the interior of the tube is lined with hydrophobic residues (left) and can accommodate a menaquinone molecule (right). **i**, A model for the reduction of menaquinone by Huc, using electrons derived from the oxidation of atmospheric $H_2$.

Source data

However, despite this strong evidence that Huc directly donates electrons from $H_2$ oxidation to respiratory quinone, Huc is not an integral membrane protein, and thus the highly hydrophobic, membrane-bound menaquinone must be transported to its substrate-binding site 94 Å away from the membrane surface. Analysis of the structure of the Huc oligomer indicates a possible mechanism by which this could occur. The interior surface of the internal cavity of Huc is lined by hydrophobic sidechains of HucS and HucM, which facilitate the binding of menaquinone and other hydrophobic molecules and has dimensions capable of accommodating menaquinone (Fig. 4f–h, Extended Data Fig. 4h and Supplementary Video 10). The conduit for menaquinone entry to and exit from the hydrophobic chamber is

probably provided by the membrane-associated Huc stalk. Overall, menaquinone molecules probably diffuse through this tube into the Huc hydrophobic chamber, where they bind the substrate acceptor site and are reduced to menaquinol before diffusing back into the membrane, where they are oxidized by the cytochrome *bcc-aa$_3$* oxidase terminal oxidase to generate proton motive force[26] (Fig. 4i). Among hydrogenases, this mechanism of quinone reduction is unique to Huc, with the other structurally characterized respiratory hydrogenases reducing quinone in the membrane via direct electron transfer to an integral membrane cytochrome *b* subunit[28]. There are, however, some similarities with the reduction mechanism of respiratory chain complex I, which removes quinone approximately 30 Å from the membrane to position it for reduction by a [4Fe–4S] cluster within its soluble arm; however, in complex I[47,48], this extraction is achieved via an enclosed channel rather than the broader hydrophobic tube and chamber observed in Huc[47].

# Conclusions

To use the trace quantities of $H_2$ present in air, the [NiFe] hydrogenases of trace-gas-scavenging bacteria require distinct properties compared with their counterparts that function under $H_2$-rich hypoxic or anoxic conditions[4,35]. Using biochemical and electrochemical characterization of Huc from the aerobic bacterium *M. smegmatis*, we show that the required properties of $O_2$ insensitivity and high affinity for $H_2$ are inherent to this hydrogenase, rather than resulting from coupling to other processes inside the bacterial cell. Moreover, through determination of the cryo-EM structure of Huc, molecular dynamics simulations, and FTIR and EPR spectroscopy, we provide strong evidence that at least partial exclusion of $O_2$ from the active site contributes to the $O_2$ insensitivity of the enzyme. Our data show that Huc has a large electrochemical overpotential that makes it uniquely tuned for the oxidation of trace quantities of $H_2$ and for the direct donation of the resulting electrons to the respiratory cofactor menaquinone. Of note, we demonstrate that Huc accesses menaquinone via a unique and highly unusual mechanism. Through the scaffolding protein HucM, the Huc complex can extract menaquinone from the membrane and transport it 94 Å to the

electron acceptor site of the enzyme. This finding greatly expands our knowledge of the possibilities for performing respiratory quinone reduction. More broadly, these findings open pathways for biocatalyst development given that all hydrogenases applied so far in whole-cell and purified enzyme systems are low-affinity enzymes that become inhibited by $O_2$. Huc, as an oxygen-insensitive high-affinity enzyme and a group 2 [NiFe] hydrogenase, provides a basis for the development of biocatalysts that operate under ambient conditions.

# Methods

## Statistics and reproducibility

### Huc purification and PAGE analysis

Huc was purified from *M. smegmatis* cells on eight separate occasions throughout the course of the study. The yield of purified Huc (mg per litre of culture) varied by a factor of five between purifications. The relative abundance of the Huc oligomer and low molecular weight species (Extended Data Fig. 1a) varied approximately twofold between purifications. SDS–PAGE and activity staining was performed on Huc from each purification ($n = 8$). The abundance of HucS, HucL and HucM subunits in the Huc oligomer (Extended Data Fig. 1b) and its activity (Extended Data Fig. 1c) was consistent between purifications, so gels from a single purification were shown. Huc oligomer from different purifications was utilized for downstream enzymatic, electrochemical and spectroscopic analysis. Statistical methods were not used to predetermine sample size for this study. Blinding and randomization were not used as they are not appropriate for this kind of study.

### HucM mutant analysis

Native PAGE analysis of Huc activity in wild-type and *ΔhucM* cell lysates was performed three times using cell lysates derived from independently grown cultures ($n = 3$ biological replicates). All replicates displayed

comparable relative activity and band sizes. A single representative gel is shown (Extended Data Fig. 1e).

### Huc differential scanning fluorimetry

Differential scanning fluorimetry experiments to determine the Huc melting temperature were performed three times on a single batch of purified Huc ($n = 3$ technical replicates) with consistent results between samples. Given the consistent purity and activity of the Huc oligomer, analysis of a single sample was sufficient to determine the melting temperature of the enzyme. Representative curves from a single experiment are shown (Extended Data Fig. 1d).

### Gas chromatography

Huc $H_2$ consumption experiments were performed on 4 different Huc preparations with each electron acceptor (NBT or menadione) ($n = 4$, biological replicates) with 3 replicate vials used in each case ($n = 3$ technical replicates). The data obtained were consistent between both biological and technical replicates. *E. coli* Hyd1 $H_2$ consumption experiments were performed on a single enzyme preparation (n = 1 biological replicates), with 3 replicate vials used ($n = 3$ technical replicates). Figure 1a,b and Extended Data Fig. 1f show data from a single enzyme preparation with data collected concurrently for all samples presented. Data are presented as a mean of the 3 technical replicates ($n = 3$) with error bars showing s.d. Control vials containing no enzyme showed no $H_2$ consumption or loss and were included in each experiment with 2 or 3 technical replicates ($n = 2$ or 3).

### Hydrogen electrode experiments

Huc oxygen tolerance and electron acceptor dependence experiments were performed 3 times on separate Huc purifications ($n = 3$, biological replicates) with 3 separate electrode traces collected for each $O_2$ concentration or electron acceptor for each Huc sample ($n = 3$, technical replicates). For menadione activity experiments, 5 technical replicates were performed due to increased instrument noise when using this compound ($n = 5$, technical

replicates). The data obtained were consistent between both biological and technical replicates. The mean of the three replicates for each condition (five for menadione) from a single Huc sample is shown with error bars showing s.d. (Fig. 1c,d).

## Protein film electrochemistry

Cyclic voltammograms of Huc in 100% $H_2$ (Fig. 1e) were recorded using a single enzyme preparation (*n* = 1 biological replicates). Two immobilized Huc films were recorded (*n* = 2 technical replicates), with each film scanned at least two times. The data obtained were consistent between technical replicates. Two control films with no immobilized enzyme were included in each experiment (*n* = 2 technical replicates), each scanned two times. Cyclic voltammetry for Huc, Hyd1 and Hyd2 in 10 ppm to 5% $H_2$ were performed on a single preparation of each enzyme (*n* = 1, biological replicates) with 3 different protein films of each (*n* = 3 technical replicates) (Fig. 1f and Extended Data Fig. 2). The Huc used in these experiments was from a different purification to that in the 100% $H_2$ experiments. It is common in protein film electrochemistry to find variation in absolute enzyme coverage between different films arising from minor differences in film preparation; therefore it was not reasonable to compare the absolute current from electrochemistry between different protein films. However, the trends in current in response to changes in $H_2$ partial pressure were consistent between technical replicates. For determination of onset potentials for Huc, the mean of three objective readings of onset potential for two Huc protein films was determined, as shown with error bars showing s.d. (Extended Data Fig. 2a).

## Cryo-EM imaging

Cryo-EM imaging was performed on Huc from 2 separate purifications (*n* = 2 biological replicates). In total, 2,226 and 3,113 images were collected from each of the purifications with single representative images of free and membrane-associated Huc down in Extended Data Fig. 3a. Two-dimensional and three-dimensional reconstructions of were performed independently on each dataset.

## EPR experiments

EPR experiments were performed on a single enzyme preparation (*n* = 1 biological replicates), incubated under either $H_2$ or Ar before flash freezing. Each individual power and temperature setting was recorded once (*n* = 1 technical replicate), and each spectrum shown represents the average of 4 scans (Fig. 3c, Extended Data Fig. 5b–d). Given the consistency of the power and temperature dependence when cross-correlated, this was sufficient for the conclusions drawn in the manuscript.

## FTIR experiments

FTIR experiments were performed on a single enzyme preparation (*n* = 1 biological replicates). The population of reduced states upon $H_2$ exposure and the formation of Ni-B upon exposure to air was performed with at least three individual attenuated total reflectance (ATR)-FTIR spectroscopy sample preparations (*n* = 3, technical replicates) leading to similar results (Fig. 3d and Extended Data Fig. 8a–d,h). The kinetics transferring between these redox state populations were recorded on one sample preparation (*n* = 1 technical replicates) (Extended Data Fig. 8e).

## Mass spectrometry

The identification of dihydromenaquinone-9 in the purified Huc protein by LC–MS (Fig. 4e) was performed once (*n* = 1 technical replicate) due to the qualitative nature of the analysis and sample quantity constraints. While the estimate of occupancy (Supplementary Fig. 5) using these data is quantitative, the purpose of that analysis was predominantly to demonstrate that the signal was consistent with that anticipated for the ligand, which only required a single replicate.

## Molecular dynamics experiments

Molecular dynamics simulations were performed 3 times for each condition simulated (*n* = 3). Cumulative gas occupancy plots are presented as the mean

of these simulations with dotted lines showing s.d. (Extended Data Figs. 6b and 7b).

**Publicly available data**

Publicly available structural coordinates utilized in this study can be accessed via the following PDB accession codes, 5AA5, 2FRV, 4U9H, 6EHQ, 4C3O, 3AYX, 5MDK.

## Bacterial strains and growth conditions

*M. smegmatis* mc$^2$155 and its derivatives were grown on lysogeny broth (LB) agar plates supplemented with 0.05% (w/v) Tween 80 (LBT). In broth culture, *M. smegmatis* was grown in either LBT, or Hartmans de Bont minimal medium supplemented with 0.2% (w/v) glycerol, 0.05% (w/v) tyloxapol, and 10 mm $NiSO_4$ (HdB). *E. coli* was maintained on LB agar plates and grown in LB broth. Selective LB or LBT medium used for cloning experiments contained gentamycin at 5 µg ml$^{-1}$ for *M. smegmatis* and 20 µg ml$^{-1}$ for *E. coli*. Cultures were incubated at 37 °C with rotary shaking at 150 rpm for liquid cultures unless otherwise specified. The strains of *M. smegmatis* and its derivatives and *E. coli* are listed in Supplementary Table 4.

## Genetic manipulation of *M. smegmatis*

For Huc production, protein expression was performed using the *M. smegmatis* mc$^2$155 strain PRC1, which lacks the glycerol response regulator *gylR* (*gylR* Leu154→frameshift) and the Hhy hydrogenase (*hhyS* inactivation)[49,50]. To facilitate Huc purification, a 2×StrepII tag was inserted at the N-terminal end of the small subunit (MSMEG_2262, HucS) through allelic exchange mutagenesis as described previously[51]. An allelic exchange construct, hucS-2×StrepII (2656 bp), was synthesized by Genewiz and cloned into the SpeI site of the mycobacterial shuttle plasmid pX33 to yield the construct pHuc-2×StrepII (Supplementary Table 4). pHuc-2×StrepII was propagated in *E. coli* DH5α and transformed into wild-type *M. smegmatis* mc$^2$155 PRC1 cells by electroporation. Gentamycin was used in selective

solid and liquid medium to propagate pX33. To allow permissive temperature-sensitive vector replication, transformants were incubated on LBT gentamycin plates at 28 °C until colonies were visible (5–7 days). The resultant catechol-positive colonies were subcultured in LBT gentamycin broth at 40 °C for 3–5 days and then diluted from broth onto fresh LBT gentamycin plates and incubated at 37 °C for 3–5 days to facilitate the integration of the recombinant plasmid, via flanking regions, into the chromosome. The second recombination event was facilitated by subculturing catechol-reactive and gentamycin-resistant colonies in LBT supplemented with 10% sucrose (w/v) for 3–5 days, followed by dilution onto LBT agar plates supplemented with 10% sucrose (w/v) and incubating at 37 °C for 3–5 days. Gentamycin-sensitive and catechol-unreactive colonies were subsequently screened by PCR to distinguish wild-type revertants from HucS-2×StrII mutants. A colony that was PCR positive for the HucS-2×StrII insertion was designated *M. smegmatis* mc$^2$155 PRC1 HucS-2×StrII and used for subsequent experiments. The primers used for screening are listed in Supplementary Table 4. HucM was deleted using allelic exchange mutagenesis using the same PRC1 parent strain and procedure for the 2×StrepII tagging of HucS. An allelic exchange construct consisting of ~1,000 bp of the genomic sequence on either side of HucM was used.

## Huc purification

*M. smegmatis* mc$^2$155 PRC1 HucS-2×StrII cells were propagated in large volumes (10–20 l) of HdB until 24 h into stationary phase. Cells were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris, 150 mM NaCl pH 8.0) supplemented with 0.1 mg ml$^{-1}$ lysozyme, 0.1 mg ml$^{-1}$ DNase, 1 mM $MgCl_2$ and protease inhibitor cocktail tablets. Cells were lysed by cell disruptor (Emulsiflex C-5), and cell debris were removed by centrifugation at 30,000*g* for 15 min. Biotin blocking buffer (IBA) was added to clarified cell lysates (1 ml per 25 ml of lysate) before loading onto a 1 ml StrepTrap column (Cytiva), and the column was washed extensively with lysis buffer before bound Huc was eluted with lysis buffer + 2.5 mM desthiobiotin. Huc-containing fractions from StrepTrap were determined by SDS–PAGE, pooled, and concentrated using a 100 kDa MWCO centrifugal concentrator (Amicon, Millipore). Concentrated Huc was loaded on a

Superose 6 10/300 column (Cytiva), fractions containing oligomeric Huc were confirmed by SDS- and native PAGE, pooled, and concentrated to ~6 mg ml$^{-1}$ before flash freezing in liquid $N_2$ and storage at −80 °C. Typically purification yielded 10–20 µg l$^{-1}$ of the Huc complex.

## PAGE analysis, activity staining and western blotting

For SDS–PAGE and native PAGE, samples were run on Bolt 4–12% SDS–PAGE and native PAGE 4-16% gels (Invitrogen) respectively, according to the manufacturer's instructions. Gels were stained for total protein by AcquaStain Protein Gel Stain (Bulldog) and hydrogenase activity using the colorimetric electron acceptor NBT hydrogenase. For NBT activity staining, gels were incubated in 50 mM Tris, 150 mM NaCl pH 8.0 supplemented with 500 µM NBT in an anaerobic jar amended with 7% $H_2$ anaerobic mix. Incubation was conducted for 30 mins to 12 h, depending on the level of activity. Bands exhibiting hydrogenase activity were identified based on the purple colour of reduced NBT. For western blotting, proteins were transferred to nitrocellulose membrane from SDS–PAGE or native PAGE gels using the Transblot Turbo semidry transfer apparatus (BioRad) at 25 V for 30 min. Following transfer, the protein-containing nitrocellulose membrane was blocked with 3% (w/v) BSA in PBS (pH 7.4) with 0.1% (v/v) Tween 20 (PBST). The nitrocellulose membrane was washed three times in 20 ml of PBST and finally resuspended in 10 ml of the same buffer. Strep-Tactin HRP conjugate was then added at a 1:100,000 dilution. Bound Strep-Tactin HRP was detected by chemiluminescence using the ECL Prime detection kit (Cytiva) following the manufacturer's specifications, blots were visualized using a ChemiDoc (BioRad).

## Amperometric measurement of Huc activity

$H_2$ consumption by purified Huc was measured amperometrically with a Unisense $H_2$ microsensor polarized to +100 mV for 1 h. Initially, the electrode was calibrated with known $H_2$ standards of 0%, 1%, and 10% $H_2$ (v/v) in gas saturated buffer (50 mM Tris, 150 mM NaCl, pH 8.0). This buffer was prepared by firstly bubbling all buffers with 100% nitrogen gas for 1 h to remove any traces of oxygen, then bubbling 100% $H_2$ (v/v)

through the de-oxygenated buffer for 10 min. Before degassing and regassing, all buffers contained electron acceptors to a final concentration of 200 µM (menadione, NBT, or benzyl viologen). For each reading, 1 ml of 1% (8 µM) $H_2$-infused buffer containing the electron acceptor was added to the microrespiration assay chamber, to a final concentration of 1% $H_2$. The electrode was subsequently placed into the chamber to equilibrate. Once equilibrated (approximately 10 min), BSA (0.3 mg $ml^{-1}$) and Huc (1–3 nM in 0.3 mg $ml^{-1}$ BSA) were added to the chamber via a needle so as not to disrupt the gas balance in the solution. The changes in $H_2$ concentration were measured using the SensorTrace Suite v3.4.00 and linear rates of $H_2$ consumption were measured from the addition of Huc until complete $H_2$ consumption. The linear rate of $H_2$ consumption by Huc was calculated between a concentration of approximately 2.5 µM $H_2$ to 0.0125 µM $H_2$ over eight time points. These rates of $H_2$ consumption were used to plot a Michaelis-Menton curve and calculate Huc-specific activity in the presence of different electron acceptors, as well as the maximal velocity of Huc.

## Mass spectrometric menaquinone detection

Samples were prepared for LC–MS analysis using a modified Folch extraction. In brief, a 100 µl solution of purified protein (~57 µg) was treated with 2000 µl of 2:1 chloroform:methanol (v/v) after which the mixture was shaken for 10 min and allowed to stand for a further 50 min. Four-hundred microlitres of water was added and the mixture was shaken for 10 min after which the sample was allowed to stand until the two phases had completely separated. The lower chloroform-rich phase was then transferred to a 2 ml sample vial and the solvent was removed under a stream of nitrogen gas. The resulting residue was reconstituted in 100 µl of 2:1 chloroform:methanol v/v and transferred to a 200 µl sample insert, the solvent was again removed and the sample reconstituted in a 7:3 mixture of LC solvent A:LC solvent B (v/v). Samples were analysed using a Dionex RSLC3000 UHPLC (Thermo) coupled to a Q-Exactive Plus Orbitrap MS (Thermo) using a C18 column (Zorbax Eclipse Plus C18 Rapid Resolution HD 2.1 × 100 mm × 1.8 µm, Agilent) with a binary solvent system; solvent A = 40% isopropanol and solvent B = 98% isopropanol both containing 2 mM formic acid and 8 mM ammonium formate. Linear gradient time, %B as

follows: 0 min, 0%; 8 min, 35%; 16 min, 50%; 19 min, 80%; 23 min, 100%; 28 min, 100%; 30 min, 0%; 32 min, 0%. The flow rate was 250 µl $min^{-1}$, the column temperature 50 °C, and the sample injection volume was 10 µl. The mass spectrometer operated at a resolution of 70,000 in polarity switching mode with the following electrospray ionization source conditions: spray voltage 3.5 kV; capillary temperature 300 °C; sheath gas 34; Aux gas 13; sweep gas 1 and probe temp 120 °C.

The occupancy of MQ9(II-$H_2$) associated with Huc was calculated by comparison against a single sample containing the standard, MQ9. MQ9 was the closest structural analogue to MQ9(II-$H_2$) (differing only in the presence of an additional double bond in the second isoprenyl unit) that was commercially available and was purchased from Toronto Research Chemicals. The peak area for the ammonium adduct of MQ9 in the standard sample containing 452.64 ng of MQ9 was $7.40 \times 10^9$ and the peak area for the ammonium adduct of MQ9(II-$H_2$) from the protein sample was $6.77 \times 10^9$ (Supplementary Fig. 5). Assuming that the response of MQ9(II-$H_2$) and MQ9 standard are equivalent, this provides an estimate of the amount of MQ9(II-$H_2$) in the protein sample of 414 ng. The protein sample contained 57 µg of Huc which corresponds to 68.4 pmoles (MW of Huc is 833 kDa). As there are eight MQ9(II-$H_2$) per Huc complex at full occupancy, this corresponds to 547.2 pmoles or 430.4 ng (MW of MQ9(II-$H_2$) = 786.63 g $mol^{-1}$) of MQ9(II-$H_2$). Consequently, the occupancy of Huc in MQ9(II-$H_2$) is estimated to be 96.2% (414 ng / 430.4 ng). If MQ9(II-$H_2$) is bound outside the observed binding sites in the Huc structure, then the actual occupancy in these sites will be lower. However, given the high quality of the density of the Huc head group present, we expect that the MQ9(II-$H_2$) present at these sites accounts for the majority in the sample.

## Mass spectrometric identification of HucM

The 18-kDa band on SDS–PAGE originating from the Huc complex was excised and extracted from the gel matrix. Upon destaining, the proteins were reduced with tris(2-carboxyethyl)phosphine (Pierce), alkylated with iodoacetamide (Sigma), and digested with mass spectrometry grade trypsin

(Promega). The extracted peptides were analysed by LC–MS/MS on an Ultimate 3000 RSLCnano System (Dionex) coupled to an Orbitrap Fusion Tribrid (ThermoFisher Scientific) mass spectrometer equipped with a nanospray source. The peptides were first loaded and desalted on an Acclaim PepMap trap column (0.1 mm id × 20 mm, 5 µm) and then separated on an Acclaim PepMap analytical column (75 µm id × 50 cm, 2 µm) over a 30 min linear gradient of 4–36% acetonitrile/0.1% formic acid. The Orbitrap Fusion Tribrid was operated in data-dependent acquisition mode with a fixed cycle time of 2 s. The Orbitrap full ms1 scan was set to survey a mass range of 375–1,800 *m*/*z* with a resolution of 120,000 at *m*/*z* 400, an AGC target of $1 \times 10^6$, and a maximum injection time of 110 ms. Individual precursor ions were selected for HCD fragmentation (collision energy 32%) and subsequent fragment ms2 scan were acquired in the Orbitrap using a resolution of 60,000 at *m*/*z* 400, an AGC target of $5 \times 10^{5,}$ and a maximum injection time of 118 ms. The dynamic exclusion was set at ±10 ppm for 10 s after one occurrence. Raw data were processed using Byonic (ProteinMetrics v4.3) against a protein database covering *M. smegmatis* mc$^2$155. The precursor mass tolerance was set at 20 ppm, and fragment ions were detected at 0.6 Da. Oxidation (*M*) was set as dynamic modification, carbamidomethyl (*C*) as fixed modification. Only peptides and proteins falling below a false discovery rate of 0.01 were reported.

## Gas chromatography analysis

Gas chromatography experiments were used to measure the consumption of $H_2$ gas by both pure Huc protein and cultures of *M. smegmatis* expressing Huc. For pure protein experiments, to assess the rate of $H_2$ oxidation of Huc with menadione as the electron acceptor, 5 ml of buffer (50 mM Tris, 150 mM NaCl, pH 8.0, 0.3 mg ml$^{-1}$ BSA) with 200 µM menadione was contained in a 120 ml sealed serum vial, in triplicate. The headspace of the vial was flushed with 3 ppm $H_2$ in an atmospheric gas mix for 10 mins and then 100 nM Huc was added via a syringe to begin the reaction. For experiments comparing the $H_2$ oxidation rate of Huc and Hyd1, the same reaction buffer contained 200 µM NBT in place of menadione and a headspace of either 10 or 100 ppm $H_2$ in $N_2$ was used, with 50 nM Huc or Hyd1 added to start the reaction. Before the experiment, Hyd1 was activated

by incubation in an atmosphere of 7% $H_2$ in $N_2$ for 18 h, before being added anaerobically to the serum vials. Vials were incubated at room temperature, stirring, and hydrogen gas concentration was monitored using a pulsed discharge helium ionization detector (model TGA-6791-W-4U-2, Valco Instruments Company) over specific time intervals, until significant $H_2$ oxidation was no longer observed, or the limit of detection had been reached.

## Differential scanning fluorimetry

To determine the stability of the Huc complex, thermal melting was performed from 20 to 90 °C using a PrometheusNT.48 DSF (Nanotemper) using high sensitivity capillaries. The ratio of change in fluorescence at 330 and 350 nm was monitored to determine protein unfolding. Melting was performed with a Huc concentration of 0.2 mg $ml^{-1}$ in buffer containing 50 mM Tris, 150 mM NaCl pH 8.0.

## Cryo-EM imaging

Samples (3 µl) were applied onto a glow-discharged UltrAuFoil grid (Quantifoil GmbH) and were flash-frozen in liquid ethane using the Vitrobot mark IV (ThermoFisher Scientific) set at 100% humidity and 4 °C for the prep chamber. Data were collected on a Titan Krios microscope (ThermoFisher Scientific) operated at an accelerating voltage of 300 kV with a 50 µm C2 aperture at an indicated magnification of 105 kx in nanoprobe EFTEM mode. Gatan K3 direct electron detector positioned post a Gatan Quantum energy filter, operated in a zero-energy-loss mode with a slit width of 10 eV was used to acquire dose fractionated images of the Huc complex without an objective aperture. Two initial Huc datasets were collected, one low concentration dataset (0.4 mg $ml^{-1}$) composed of 2,226 movies, and a medium concentration dataset (1 mg $ml^{-1}$) composed of 3,113 movies. Movies were recorded in hardware-binned mode (previously called counted mode on the K3 camera) yielding a physical pixel size of 0.82 Å $pixel^{-1}$ with an exposure time of 6 s amounting to a total dose of 66.0 $e^-$ $Å^{-2}$ at a dose rate of ~7.5 $e^-$ $pixel^{-1}$ $s^{-1}$, which was fractionated into 60 subframes. A further high concentration dataset (4 mg $ml^{-1}$) of 9,868 micrographs was

also recorded using the same microscope but at a higher magnification of 165 kx on K3 detector, the resulting pixel size was 0.5 Å. The slit width for zero-loss filtering was set to 10 eV and images were collected at a dose rate of 4.37 $e^-$ pixel $s^{-1}$. Movies were collected with an exposure time of 4 s amounting to a total dose of 60.4 $e^-$ $Å^{-2}$, which were further fractionated into 60 subframes. The defocus range was set between −1.3 and −0.3 μm.

## Cryo-EM data processing and analysis

Micrographs from all datasets were motion-corrected using UCSF Motioncor 1.0.4 and dose-weighted averages had their contrast transfer function (CTF) parameters estimated using CTFFIND 4.1.8, implemented using Relion 3.1.2 (ref. [52]).

### Huc oligomer reconstruction from dataset 1

Particle coordinates were determined by crYOLO 1.7.6 using a model trained on manually particles picked from 20 micrographs[53]. Unbinned particles were extracted from micrographs using Relion 3.1.2, before being imported into cryoSPARC 3.3.1 for initial 2D classification to remove bad particles, followed by ab initio model generation and 3D refinement[54]. Refined particles were reimported into Relion 3.1.2 and CTF refinement was performed, followed by Bayesian polishing[52]. Particles were reimported into cryoSPARC 3.3.1 for final 2D classification to remove residual bad particles, followed by non-uniform 3D consensus refinement to generate final maps at 2.19 Å (Fourier shell correlation (FSC) = 0.143, gold standard).

### Huc$S_2L_2$ subunit dataset 1

Huc$S_2L_2$ particles were picked using gautomatch V 0.53 (developed by K. Zhang) using a diameter of 80 Å and a minimum distance between picked particles set to 20 Å. The resultant particles were extracted and binned 4 times to a box size of 56 pixels. A small subset of unbinned particles with 224 pixel box size were subjected to 2D classification to exclude noisy particles and the remainder were selected and subjected to 3D ab initio model generation with 4 classes using cryoSPARC v3.0.1 (ref. [54]). A class

corresponding to a population of $HucS_2L_2$ subunits showed clear $C_2$ symmetry and this particle set was subjected to homogeneous refinement with $C_2$ symmetry applied resulting in a 3.25 Å resolution reconstruction. The resulting volume was used as the initial model for further classification and refinement of the full dataset as shown in the workflow Supplementary Fig. 3. Particles corresponding to the subunits were selected from binned dataset following 2D classification and heterogeneous refinement. The cryoSPARC 3D alignments from heterogeneous refinement were then exported to Relion using the csparc2star.py script in the pyem v0.5 package. In Relion, particles were re-extracted with no binning using the refined coordinates from the imported particles. These were re-imported into cryoSPARC and subjected to heterogeneous refinement followed by homogeneous refinement with $C_2$ symmetry applied. The FSC from the resultant reconstruction showed particle duplication as a result of over-picking using gautomatch. The refined particle coordinates were imported into Relion as described above and were cleaned to remove duplicated particles. The cleaned re-extracted particles were subjected to a round of CTF refinement followed by auto-refinement and Bayesian polishing. The resultant polished particles were imported back into cryoSPARC and CTF refinement, and non-uniform refinement with CTF parameter optimization was performed to reach a final global resolution of 1.67 Å (FSC = 0.143, gold standard), which was close to the Nyquist limit for the dataset (1.64 Å).

**$HucS_2L_2$ subunit dataset 2**

A second dataset at higher magnification was collected to overcome the Nyquist limit imposed on the map resolution of dataset 1. The particles were picked, extracted, and binned as described above and in Supplementary Fig. 2. Multiple rounds of 2D classification, ab initio classification, and heterogeneous refinement was performed to retain only particles containing the full Huc oligomer. Homogeneous refinement was performed on the binned particles with the $HucS_2L_2$ subunit map as an initial model in order to centre the particles on the $HucS_2L_2$ subunit. These particles were then re-extracted in Relion based on the refined coordinates to a box size of 380 pixels. Homogeneous refinement with $C_2$ symmetry resulted in a subunit reconstruction map of 2.15 Å. This step ensured the retention of all $HucS_2L_2$

particles with a good signal to noise ratio. In order to ensure that all symmetry-related $HucS_2L_2$ particles corresponding to the Huc oligomer are retained, the above-selected particles were re-extracted with a larger box size of 512 pixels and rescaled to 128 pixels (4× binned). The resultant particle set was subjected ab initio model generation followed by homogeneous refinement using $C_4$ symmetry to yield a 4.10 Å map of the Huc oligomer. The particles were re-extracted based on the refined coordinates from the previous step at a box size of 576 pixels and then rescaled to 288 pixels (2× binned). These particles were then subjected to homogeneous refinement followed by CTF refinement and Non-Uniform refinement in cryoSPARC to yield a 2.05 Å resolution map[54]. The refined particle set was then re-imported into Relion, and Bayesian polishing was performed. The resultant 'shiny' particles were extracted to a final box size of 576 pixels with no binning and were re-imported into cryoSPARC to perform symmetry expansion and final refinement. Duplicate particles were excluded to retain 153,359 Huc oligomers. Resultant particles were subjected to one round of refinement imposing $C_4$ symmetry, which yielded a 2.19 Å map. The resultant refined particles were symmetry expanded and subjected to local refinement within cryoSPARC 3.3.2 with no symmetry to yield a 2.16 Å map for $HucS_2L_2$. The resultant vol and particles were aligned to $C_2$ symmetry axis to exploit the $C_2$ symmetry within the $HucS_2L_2$. The symmetry aligned particles were then refined with $C_2$ symmetry followed by iterative Global CTF refinement, per particle defocus refinement followed by Ewald Sphere correction and a final round of per particle defocus refinement to reach a final global resolution of 1.52 Å (FSC = 0.143, gold standard).

### HucM stalk reconstruction

The particle set resulting in a 2.15 Å map during the $HucS_2L_2$ subunit reconstruction from dataset 2 was re-extracted at a box size of 512 pixels and rescaled to 128 pixels (4× binned). The resultant particle set was subjected to 2D classification followed by ab initio model generation and homogeneous refinement using $C_4$ symmetry to yield a 4.10 Å map of Huc oligomer. The particles were re-extracted based on the refined coordinates from the previous step at a box size of 576 pixels and then rescaled to 288

pixels (2× binned). These particles were then subjected to homogeneous refinement followed by CTF refinement and non-uniform refinement in cryoSPARC to yield 2.05 Å resolution map[54]. The refined particles set was then re-imported into Relion and Bayesian polishing was performed, the resultant 'shiny' particles were extracted to a final box size of 576 pixels with no binning. These particles were subjected to ab initio model generation with three classes to retain particles that showed clear density for the HucM tail region. The selected particles were then ctf refined and Non-Uniform refined to yield a 2.09 Å map. The stalk density was isolated using volume tools in ChimeraX 1.3 and low pass filtered to 20 Å to derive the initial model for local refinement in cryosparc[55]. A dilated cosine padded softmask was generated using the initial model. The coordinates at the attachment of the stalk to the main body of the subunit were determined using the volume tracer tool in Chimera and used to define new fulcrum point for localized refinement (Supplementary Fig. 4). Two rounds of masked local refinement resulted in improving the density of the stalk region associated with the membrane, as shown in the figure to a resolution of 5.7 Å (no mask). The resultant FSC calculates a map resolution of 5.21 Å (FSC = 0.143, gold standard) but the map features are more consistent with a resolution of 6–8 Å on visual inspection.

## Huc model building and visualization

A model of the HucS and HucL subunit dimers was generated using AlphaFold and docked into one-half of the high-resolution Huc Dimer maps using ChimeraX 1.3 (refs. [56,57]). The model was refined and rebuilt into map density using Coot 0.92 (ref. [58]). [NiFe], [3Fe–4S], and menaquinone cofactors associated with Huc were downloaded from the PDB and customized restraints generated using Elbow within the PHENIX package before they were fitted and refined into maps using Coot[59,60]. The model was then refined using real-space refinement within PHENIX[61]. Once model building was complete the model was symmetry expanded using the Map symmetry tool and waters were added using DOUSE within the PHENIX package[60]. The refined dimer model was docked into the Huc oligomer map using ChimeraX[56], with HucM manually built into the available density for one subunit using Coot[58], followed by iterative real-space refinement within PHENIX and further model building. Once model building was complete the

model was symmetry expanded using the Map symmetry tool within the PHENIX package[60]. Model quality was validated using MolProbity[62]. Images and movies were generated in Coot, ChimeraX, and Pymol[56,58]. The location and diameter of the gas channels of Huc and other [NiFe] hydrogenases was determined using the CAVER code[63].

## AlphaFold structural modelling

AlphaFold modelling was performed using AlphaFold version 2.1.1 implemented on the MASSIVE M3 computing cluster[57,64]. The sequence of the HucM C-terminal region (amino acids 80-189) was provided and modelling was run in multimer mode, with four molecules of HucM requested. The five ranked models produced by AlphaFold were compared for consistency with the top-ranked model used for further analysis and figure generation.

## Molecular dynamics parameterization and simulation

The 1.52 Å cryo-EM structure of Huc hydrogenase was used to construct the simulation system using the Charmm36m forcefield[65,66] with the TIP3P[67] water model. Sodium counterions were added to the solution to neutralize the system charge. Three sets of simulations were set up containing no gas molecules or containing 250 hydrogen or 250 oxygen molecules enriched in solution. Gas molecules were randomly inserted into the water phase. Each simulation was initiated with different starting velocities and gas positions and three independent simulations of each system were performed. Simulations were run for 50 ns (wild-type) or 100 ns (mutants). Molecular visualization and analyses were performed using VMD software[68]. The topologies of the metal cofactors were constructed based on the coordinates from the cryo-EM structure. Previous work on hydrogenase cofactors[69] has demonstrated the transferability of structurally identical metallic cofactors between hydrogenases. Thus, atomic partial charges and force constants for the bond lengths of the metal cofactors were taken from previous studies of the structurally similar *Desulfovibrio fructosovorans* [NiFe] hydrogenase[69]. Equilibrium values of bond lengths and angles within the metal cofactor were taken directly from the cryo-EM structure. Molecular dynamics

simulations were performed in a single redox state. The ready active Ni-SI form of the active site and oxidized form of the iron–sulfur clusters were chosen based on previous similar analyses[70]. The [NiFe] cofactor had bond constraints placed across the CN and CO ligands as well as between the iron atom and two bound sulfur atoms. Lennard–Jones parameters were taken from the Charmm36m[65,66] forcefield where appropriate, with exception of the cyanuric carbon[71] with the iron parameters taken from physical data to have a non-zero epsilon value[72]. The QL molecular oxygen model developed by Javanainen et al.[73] was chosen to better define the quadrupole within the molecule. The oxygen atoms are given mass and non-zero Lennard–Jones parameters as well as a partial charge. A massless virtual site, OD1, exists at the midpoint of two oxygen atoms to carry the neutralizing charge, simulating the charge distribution within the oxygen quadrupole. Likewise, the hydrogen model was given mass and charge distribution coherent with the work of Hunter et al.[70]. Molecular dynamics simulations were performed with the GROMACS v 2021.3 simulations suite. Energy minimization was performed using the steepest descent algorithm followed by a second minimization step using the conjugate gradient approach, both set to a maximum force of 250 kJ $mol^{-1}$ $nm^{-1}$. Equilibration simulations were performed within the NVT ensemble at 310 K. Position restraints were applied to the protein, cofactors and magnesium ion, with a $10^3$ kJ $mol^{-1}$ $nm^{-2}$ force constant for 100 ps. Further equilibration was carried out within the NPT ensemble with pressure maintained at 1 bar using the Parrinello–Rahman barostat[74], for 100 ps. Production simulations in the NPT ensemble were each of duration 50 ns, at a temperature of 310 K. Long-range electrostatics were treated with the Particle Mesh Ewald method[75], with values for the real-space long-range cut-off for these and Van der Waals forces being set to 10 Å. Covalent bonds containing hydrogen atoms were restrained using the LINCS algorithm[76,77] allowing a 2 fs integration step.

## EPR spectroscopy

The composition of the iron–sulfur clusters of Huc was analysed by EPR spectroscopy. X-band EPR measurements were performed on a Bruker ELEXYS E500 spectrometer equipped with a SuperX EPR049 microwave

bridge and a cylindrical $TE_{011}$ ER 4122SHQE cavity in connection with an Oxford Instruments continuous flow cryostat. Measuring temperatures were achieved using liquid helium flow through an ITC 503 temperature controller (Oxford Instruments). Samples of Huc isolated in air were prepared under a neat argon atmosphere or flash-frozen following incubation under 1 atm $H_2$, and X-band EPR spectra were collected in the temperature range 40–7 K, at varying microwave powers. Preliminary simulations of the Ni signals in the EPR spectra were carried out using EasySpin v.6.0.0-dev47 (ref. [78]).

## FTIR spectroscopy

A 4 µl volume of 5 mg $ml^{-1}$ Huc enzyme solution in a buffer containing 50 mM Tris, 150 mM NaCl pH 8 was deposited on an ATR crystal surface. The sample was applied under laboratory atmosphere (air), dried under 100% nitrogen gas, and rehydrated with a humidified aerosol (100 mM Tris-HCl (pH 8)) as described previously[79,80]. A custom-built PEEK cell (inspired by Stripp et al.[81]) that allows for gas exchange sealed the ATR unit (BioRadII from Harrick) mounted in the FTIR spectrometer (Vertex V70v, Bruker). Spectra were recorded with 2 $cm^{-1}$ resolution, 80 Hz scanner velocity, and averaged over a varying number of scans (at least 100 scans). All experiments were performed at ambient conditions (room temperature and pressure, hydrated enzyme films). Data were analysed using OPUS 8 and Origin 2021 software. To assess the response of the Huc [NiFe] cluster to ambient air, $N_2$, and $H_2$, the PEEK cell was flushed sequentially with each gas in the following sequence: $N_2$ (ambient air isolated enzyme), $^2H_2$, $N_2$, $H_2$, ambient air. Changes in the Huc spectra were allowed to stabilize before the spectra in each gas were collected.

## Protein film electrochemistry

**Huc in 100% $H_2$ experiment**

Protein film electrochemistry experiments were carried out under anaerobic conditions. The 3-electrode system was made up of (1) Ag/AgCl (4 M KCl) as reference electrode, (2) 5 mm diameter pyrolytic graphite edge (PGE) encapsulated in epoxy as the rotating disk working electrode, and (3) graphite rod as the counter electrode. The glass cell featured a water jacket for temperature control and a gas inlet/outlet for $H_2$ flow. The buffer used was composed of 5 mM 2-[*N*′-morpholino]-ethane sulfonic acid (MES), 5 mM 2-[*N*′-cyclohexylamino]ethane sulfonic acid (CHES), 5 mM *N*′-[2-hydroxyethyl]piperazine-*N*′−2- ethane sulfonic acid (HEPES), 5 mM *N*′-tris[hydroxymethyl]methyl-3-amino propane sulfonic acid (TAPS), 5 mM sodium acetate, with 0.1 M $Na_2SO_4$ as carrying electrolyte, titrated with $H_2SO_4$ to pH 7.0 at 20 °C, and purged with $N_2$ for 3 to 4 h. To remove residual $O_2$ in the PGE electrode, cyclic voltammograms were run at 100 mV $s^{-1}$ from 0 to −800 mV (versus SHE, standard hydrogen electrode) for 10 scans. To adhere Huc to the deaerated PGE electrode the surface was abraded with P1200 sandpaper and rinsed with purified water. An aliquot of 5 mg $ml^{-1}$ Huc enzyme solution in a buffer containing 50 mM Tris, 150 mM NaCl pH 8 was mixed with polymyxin B sulfate and transferred to the electrode surface. The cell solution was then saturated with $H_2$ (1 bar, 10 min) before the cyclic voltammogram of the system with the immobilized enzyme was recorded at 10 mV $s^{-1}$. The cyclic voltammogram of the blank electrode was recorded with no enzyme immobilized. Electrochemical data were acquired using a PGSTAT10 and GPES 4.9 software (Metrohm/Autolab). Data were analysed using Origin 8 software. All potential values are referenced versus SHE. Experiments were conducted on two independent Huc films, with each film scanned at least two times.

**Huc, Hyd1 and Hyd2 in 0 to 5% $H_2$ experiments**

Protein film electrochemistry experiments were conducted in a $N_2$-filled anaerobic glove box (Glove Box Technology Limited, $O_2$ < 2 ppm). Experiments were carried out using a glass electrochemical cell with gas

inlet and outlet valve[82]. The main cell compartment was surrounded by a water jacket to control temperature. All experiments were performed at 25 °C. A saturated calomel electrode (SCE) was used as the reference, and was held in an isolated glass arm containing 0.10 M NaCl and connected to the main cell compartment by a Luggin capillary. The SCE was calibrated with ferrocene-methanol as +244 mV versus SHE at 25 °C, and potentials are corrected back to *V* versus SHE in all figures. The working electrode was a PGE rotating disk electrode, 2 mm diameter, and was polished with silicon carbide paper (first P1200 and then P4000). Electrode rotation was controlled by a Metrohm Autolab IME663 rotator, and the electrode was rotated at 2,000 rpm in all experiments shown here in order to achieve effective mass transport to/from the immobilized enzyme film. A platinum wire was used as the counter electrode. Cyclic voltammetry was controlled by an Autolab PGSTAT 10 potentiostat using NOVA software version 1.10. All experiments were conducted with gas flushing through the headspace of the electrochemical cell. Gas mixtures were prepared from cylinders of pure Ar, pure $H_2$, 5% $H_2$ in $N_2$, 1,000 ppm $H_2$ in $N_2$ (all from BOC), using gas mass flow controllers (Smart-Trak2, Sierra Instruments) to obtain the desired partial pressure of $H_2$ within an inert carrier gas (Ar or $N_2$). A mixed buffer system was used in all experiments; this consisted of 15 mM in each of MES, HEPES, TAPS, CHES (all from Melford) and sodium acetate (Sigma), with 0.1 M NaCl (Fisher) as supporting electrolyte. All solutions were prepared using purified water (Millipore: resistivity 18.2 MΩ cm) and titrated with HCl to pH 7.0 at 25 °C. Buffers were flushed with $N_2$ overnight to remove $O_2$ before being taken into the glove box. For preparation of films of Huc, 2 μl enzyme solution (containing 20 mg ml$^{-1}$ polymyxin B sulfate as coadsorbate) was spotted onto a freshly polished PGE electrode surface and left for 3 min; the electrode was then rinsed with purified water to remove unabsorbed enzyme. The electrode was placed in the electrochemical cell containing 8 ml of pristine mixed buffer. To minimize protein film loss during measurements, 0.025 mg ml$^{-1}$ polymyxin B sulfate was added to the mixed buffer. The cell headspace was flushed with gas for at least 15 min to fully equilibrate with the cell solution before applying a potential. Cyclic voltammograms were recorded with constant gas flow through the headspace of the cell at 1 bar. *E. coli* Hyd1 and Hyd2 were prepared as described previously[83,84]. These hydrogenases were first

activated under 100% $H_2$ at room temperature for at least 5 h. Films of Hyd1 and Hyd2 were prepared as for Huc, but without use of polymyxin B sulfate.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

Cryo-EM maps and atomic models generated in this paper have been deposited in the Protein Data Bank (accession codes 7UTD, 7UUR, 7UUS, 8DQV) and the Electron Microscopy Data Bank (accession codes EMD-26767, EMD-26801, EMD-26802, EMD-27661). Raw data from the Huc molecular dynamics simulations are available via Zenodo (https://zenodo.org/record/7378976). Source data are provided with this paper.

## References


1. Greening, C. & Grinter, R. Microbial oxidation of atmospheric trace gases. *Nat. Rev. Microbiol.* **20**, 513–528 (2022).

2. Bay, S. K. et al. Trace gas oxidizers are widespread and active members of soil microbial communities. *Nat. Microbiol.* **6**, 246–256 (2021).

3. Ortiz, M. et al. Multiple energy sources and metabolic strategies sustain microbial diversity in Antarctic desert soils. *Proc. Natl Acad. Sci. USA* **118**, e2025322118 (2021).

4. Greening, C., Berney, M., Hards, K., Cook, G. M. & Conrad, R. A soil actinobacterium scavenges atmospheric $H_2$ using two membrane-associated, oxygen-dependent [NiFe] hydrogenases. *Proc. Natl Acad. Sci. USA* **111**, 4257–4261 (2014).

5. Greening, C. et al. Genomic and metagenomic surveys of hydrogenase distribution indicate $H_2$ is a widely utilised energy source for microbial growth and survival. *ISME J.* **10**, 761 (2016).

6. Constant, P., Chowdhury, S. P., Pratscher, J. & Conrad, R. Streptomycetes contributing to atmospheric molecular hydrogen soil uptake are widespread and encode a putative high-affinity [NiFe]-hydrogenase. *Environ. Microbiol.* **12**, 821–829 (2010).

7. Greening, C., Villas-Bôas, S. G., Robson, J. R., Berney, M. & Cook, G. M. The growth and survival of *Mycobacterium smegmatis* is enhanced by co-metabolism of atmospheric $H_2$. *PLoS ONE* **9**, e103034 (2014).

8. Koch, H. et al. Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science* **345**, 1052–1054 (2014).

9. Tveit, A. T. et al. Widespread soil bacterium that oxidizes atmospheric methane. *Proc. Natl Acad. Sci. USA* **116**, 8515–8524 (2019).

10. Islam, Z. F. et al. A widely distributed hydrogenase oxidises atmospheric $H_2$ during bacterial growth. *ISME J.* **14**, 2649–2658 (2020).

11. Schmitz, R. A. et al. The thermoacidophilic methanotroph *Methylacidiphilum fumariolicum* SolV oxidizes subatmospheric $H_2$ with a high-affinity, membrane-associated [NiFe] hydrogenase. *ISME J.* **14**, 1223–1232 (2020).

12. Islam, Z. F. et al. Two Chloroflexi classes independently evolved the ability to persist on atmospheric hydrogen and carbon monoxide. *ISME J.* **13**, 1801–1813 (2019).

13. Greening, C. et al. Persistence of the dominant soil phylum Acidobacteria by trace gas scavenging. *Proc. Natl Acad. Sci. USA* **112**, 10497–10502 (2015).

14. Liot, Q. & Constant, P. Breathing air to save energy–new insights into the ecophysiological role of high-affinity [NiFe]-hydrogenase in *Streptomyces avermitilis*. *Microbiologyopen* **5**, 47–59 (2016).

15. Ji, M. et al. Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* **552**, 400 (2017).

16. Berney, M. & Cook, G. M. Unique flexibility in energy metabolism allows mycobacteria to combat starvation and hypoxia. *PLoS ONE* **5**, e8614 (2010).

17. Berney, M., Greening, C., Hards, K., Collins, D. & Cook, G. M. Three different [NiFe] hydrogenases confer metabolic flexibility in the obligate aerobe *Mycobacterium smegmatis*. *Environ. Microbiol.* **16**, 318–330 (2014).

18. Novelli, P. C. et al. Molecular hydrogen in the troposphere: Global distribution and budget. *J. Geophys. Res.* **104**, 30427–30444 (1999).

19. Gatley, D. P., Herrmann, S. & Kretzschmar, H.-J. A twenty-first century molar mass for dry air. *HVAC&R Res.* **14**, 655–662 (2008).

20. Volbeda, A. et al. Crystal structure of the nickel–iron hydrogenase from *Desulfovibrio gigas*. *Nature* **373**, 580–587 (1995).

21. Shomura, Y., Yoon, K.-S., Nishihara, H. & Higuchi, Y. Structural basis for a [4Fe-3S] cluster in the oxygen-tolerant membrane-bound [NiFe]-hydrogenase. *Nature* **479**, 253–256 (2011).

22. Fritsch, J. et al. The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. *Nature* **479**, 249–252 (2011).

23. Schäfer, C. et al. Structure of an actinobacterial-type [NiFe]-hydrogenase reveals insight into $O_2$-tolerant $H_2$ oxidation. *Structure* **24**, 285–292 (2016).

24. Ogata, H., Nishikawa, K. & Lubitz, W. Hydrogens detected by subatomic resolution protein crystallography in a [NiFe] hydrogenase. *Nature* **520**, 571–574 (2015).

25. Leung, P. M. et al. A nitrite-oxidising bacterium constitutively consumes atmospheric hydrogen. *ISME J.* **16**, 2213–2219 (2022).

26. Cordero, P. R. et al. Two uptake hydrogenases differentially interact with the aerobic respiratory chain during mycobacterial growth and persistence. *J. Biol. Chem.* **294**, 18980–18991 (2019).

27. Lukey, M. J. et al. How *Escherichia coli* is equipped to oxidize hydrogen under different redox conditions. *J. Biol. Chem.* **285**, 3928–3938 (2010).

28. Volbeda, A. et al. Crystal structure of the $O_2$-tolerant membrane-bound hydrogenase 1 from *Escherichia coli* in complex with its cognate cytochrome b. *Structure* **21**, 184–190 (2013).

29. Pinske, C., Jaroschinsky, M., Sargent, F. & Sawers, G. Zymographic differentiation of [NiFe]-hydrogenases 1, 2 and 3 of *Escherichia coli* K-12. *BMC Microbiol.* **12**, 1–12 (2012).

30. Thauer, R. K., Kaster, A.-K., Seedorf, H., Buckel, W. & Hedderich, R. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* **6**, 579–591 (2008).

31. Wardman, P. The reduction potential of benzyl viologen: an important reference compound for oxidant/radical redox couples. *Free Radic. Res. Commun.* **14**, 57–67 (1991).

32. Shafaat, H. S., Rüdiger, O., Ogata, H. & Lubitz, W. [NiFe] hydrogenases: a common active site for hydrogen metabolism under diverse conditions. *Biochim. Biophys. Acta* **1827**, 986–1002 (2013).

33. Parkin, A. et al. How Salmonella oxidises $H_2$ under aerobic conditions. *FEBS Lett.* **586**, 536–544 (2012).

34. Vincent, K. A. et al. Electrochemical definitions of $O_2$ sensitivity and oxidative inactivation in hydrogenases. *J. Am. Chem. Soc.* **127**, 18179–18189 (2005).

35. Lubitz, W., Ogata, H., Rudiger, O. & Reijerse, E. Hydrogenases. *Chem. Rev.* **114**, 4081–4148 (2014).

36. Zanello, P. Structure and electrochemistry of proteins harboring iron-sulfur clusters of different nuclearities. Part II. [4Fe-4S] and [3Fe-4S] iron-sulfur proteins. *J. Struct. Biol.* **202**, 250–263 (2018).

37. Adamson, H. et al. Retuning the catalytic bias and overpotential of a [NiFe]-hydrogenase via a single amino acid exchange at the electron entry/exit site. *J. Am. Chem. Soc.* **139**, 10677–10686 (2017).

38. Vincent, K. A. et al. Electricity from low-level $H_2$ in still air–an ultimate test for an oxygen tolerant hydrogenase. *Chem. Commun.* **48**, 5033–5035 (2006).

39. Wait, A. F., Parkin, A., Morley, G. M., dos Santos, L. & Armstrong, F. A. Characteristics of enzyme-based hydrogen fuel cells using an oxygen-tolerant hydrogenase as the anodic catalyst. *J. Phys. Chem. C* **114**, 12003–12009 (2010).

40. Wulff, P., Day, C. C., Sargent, F. & Armstrong, F. A. How oxygen reacts with oxygen-tolerant respiratory [NiFe]-hydrogenases. *Proc. Natl Acad. Sci. USA* **111**, 6606–6611 (2014).

41. Wulff, P., Thomas, C., Sargent, F. & Armstrong, F. A. How the oxygen tolerance of a [NiFe]-hydrogenase depends on quaternary structure. *J. Biol. Inorg. Chem.* **21**, 121–134 (2016).

42. Liebgott, P.-P. et al. Relating diffusion along the substrate tunnel and oxygen sensitivity in hydrogenase. *Nat. Chem. Biol.* **6**, 63–70 (2010).

43. Abou Hamdan, A. et al. $O_2$-independent formation of the inactive states of NiFe hydrogenase. *Nat. Chem. Biol.* **9**, 15–17 (2013).

44. Fourmond, V., Infossi, P., Giudici-Orticoni, M.-T., Bertrand, P. & Léger, C. "Two-step" chronoamperometric method for studying the anaerobic inactivation of an oxygen tolerant NiFe hydrogenase. *J. Am. Chem. Soc.* **132**, 4848–4857 (2010).

45. Schäfer, C., Friedrich, B. & Lenz, O. Novel, oxygen-insensitive group 5 [NiFe]-hydrogenase in *Ralstonia eutropha*. *Appl. Environ. Microbiol.* **79**, 5137–5145 (2013).

46. Reeve, H. A. et al. Enzymes as modular catalysts for redox half-reactions in $H_2$-powered chemical synthesis: from biology to technology. *Biochem. J.* **474**, 215–230 (2017).

47. Kampjut, D. & Sazanov, L. A. The coupling mechanism of mammalian respiratory complex I. *Science* **370**, eabc4209 (2020).

48. Gutiérrez-Fernández, J. et al. Key role of quinone in the mechanism of respiratory complex I. *Nat. Commun.* **11**, 4135 (2020).

49. Cordero, P. R. *Cellular and Molecular Basis of Atmospheric Hydrogen and Carbon Monoxide Oxidation in Mycobacteria*. Thesis, Monash Univ. (2020).

50. Bong, H.-J., Ko, E.-M., Song, S.-Y., Ko, I.-J. & Oh, J.-I. Tripartite regulation of the glpFKD operon involved in glycerol catabolism by gylr, CRP, and Sigf in mycobacterium smegmatis. *J. Bacteriol.* **201**, e00511–e00519 (2019).

51. Islam, Z. F., Cordero, P. R. & Greening, C. Putative iron-sulfur proteins are required for hydrogen consumption and enhance survival of mycobacteria. *Front. Microbiol.* **10**, 2749 (2019).

52. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

53. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).

54. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

55. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

56. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

57. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

58. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

59. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

60. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

61. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).

62. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

63. Chovancova, E. et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **8**, e1002708 (2012).

64. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.04.463034 (2022).

65. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).

66. Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **44**, 129–138 (1977).

67. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

68. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).

69. Smith, D. M., Xiong, Y., Straatsma, T., Rosso, K. M. & Squier, T. C. Force-field development and molecular dynamics of [NiFe] hydrogenase. *J. Chem. Theory Comput.* **8**, 2103–2114 (2012).

70. Teixeira, V. H., Baptista, A. M. & Soares, C. M. Pathways of $H_2$ toward the active site of [NiFe]-hydrogenase. *Biophys. J.* **91**, 2035–2045 (2006).

71. Lee, M. W., Carr, J. K., Göllner, M., Hamm, P. & Meuwly, M. 2D IR spectra of cyanide in water investigated by molecular dynamics simulations. *J. Chem. Phys.* **139**, 054506 (2013).

72. Maghfiroh, C.Y., Arkundato, A., Misto & Maulina, W. In *Journal of Physics: Conference Series* https://doi.org/10.1088/1742-6596/1491/1/012022 (IOP Publishing, 2022).

73. Javanainen, M., Vattulainen, I. & Monticelli, L. On atomistic models for molecular oxygen. *J. Phys. Chem. B* **124**, 6943–6946 (2020).

74. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

75. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**,

10089–10092 (1993).

76. Hess, B., Bekker, H., Berendsen, H. J. & Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

77. Hess, B. P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).

78. Stoll, S. & Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* **178**, 42–55 (2006).

79. Senger, M. et al. Protonation/reduction dynamics at the [4Fe-4S] cluster of the hydrogen-forming cofactor in [FeFe]-hydrogenases. *Phys. Chem. Chem. Phys.* **20**, 3128–3140 (2018).

80. Senger, M. et al. Stepwise isotope editing of [FeFe]-hydrogenases exposes cofactor dynamics. *Proc. Natl Acad. Sci. USA* **113**, 8454–8459 (2016).

81. Stripp, S. T. In situ infrared spectroscopy for the analysis of gas-processing metalloenzymes. *ACS Catal.* **11**, 7845–7862 (2021).

82. Vincent, K. A., Parkin, A. & Armstrong, F. A. Investigating and exploiting the electrocatalytic properties of hydrogenases. *Chem. Rev.* **107**, 4366–4413 (2007).

83. Ramirez, M. et al. $H_2$-driven reduction of flavin by hydrogenase enables cleaner operation of nitroreductases for nitro-group to amine reductions. *Front. Catal.* https://doi.org/10.3389/fctls.2022.906694 (2022).

84. Hexter, S. V., Chung, M.-W., Vincent, K. A. & Armstrong, F. A. Unusual reaction of [NiFe]-hydrogenases with cyanide. *J. Am. Chem. Soc.* **136**, 10470–10477 (2014).

85. Senger, M., Laun, K., Soboh, B. & Stripp, S. T. Infrared characterization of the bidirectional oxygen-sensitive [NiFe]-hydrogenase from *E. coli*. *Catalysts* **8**, 530 (2018).

## Acknowledgements

We thank the Bio21 Institute for the use of the Thermo Glacios housed in their facilities during cryo-EM sample preparation and screening. We thank S. Carr for providing purified Hyd1 and Hyd2; S. Frielingsdorf, J. Rossjohn, F. Armstrong, B. Murphy and G. Knott for their review and constructive feedback on the manuscript. We acknowledge the use of instruments and assistance at the Monash Ramaciotti Centre for Cryo-Electron Microscopy, a Node of Microscopy Australia. This work was supported by an Australian Research Council (ARC) DECRA Fellowship (DE170100310) (to C.G.), an ARC Discovery Project Grant (DP200103074) (to R.G. and C.G.), a National Health and Medical Research Council Emerging Leader Grant (NHMRC) (APP1178715) (to C.G.), a NHMRC Emerging Leader Grant (APP1197376) (to R.G.), ARC LIEF grants (LE200100045, LE120100090) for the Titan Krios Gatan K3 Camera and for the Titan Krios, the Deutsche Forschungsgemeinschaft through Priority Program 1927 (1554/5-1) (to S.T.S.), a Swedish Energy Agency (grant no 48574-1) (to G.B.), a European Research Council grant (714102) (to G.B.), the European Union's Horizon 2020 research and innovation program (Marie Skłodowska Curie Grant no. 897555) (to M.S.), and a United States National Science Foundation grant (CHE-2108684) (to H.S.S.).

## Author information

Author notes

1. These authors contributed equally: Rhys Grinter, Ashleigh Kropp

### Authors and Affiliations

1. Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

   Rhys Grinter, Ashleigh Kropp & Chris Greening

2. Ramaciotti Centre for Cryo-Electron Microscopy, Monash University, Clayton, Victoria, Australia

   Hari Venugopal

3. Department of Chemistry, Ångström Laboratory, Uppsala University, Uppsala, Sweden

   Moritz Senger, Princess R. Cabotaje, Ping Huang & Gustav Berggren

4. Department of Biochemistry, University of Oxford, Oxford, UK

   Jack Badley, Ruyu Jia & Syma Khalid

5. Department of Chemistry, University of Oxford, Inorganic Chemistry Laboratory, Oxford, UK

   Zehui Duan & Kylie A. Vincent

6. Department of Physics, Experimental Molecular Biophysics, Freie Universität Berlin, Berlin, Germany

   Sven T. Stripp

7. Department of Biochemistry, Monash Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

   Christopher K. Barlow & Ralf B. Schittenhelm

8. Monash Proteomics and Metabolomics Facility, Monash Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

   Christopher K. Barlow & Ralf B. Schittenhelm

9. Centre for Electron Microscopy of Membrane Proteins, Monash Institute of Pharmaceutical Sciences, Parkville, Victoria, Australia

Matthew Belousoff

10. Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA

Hannah S. Shafaat

11. Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand

Gregory M. Cook

12. Securing Antarctica's Environmental Future, Monash University, Clayton, Victoria, Australia

Chris Greening

13. Centre to Impact AMR, Monash University, Clayton, Victoria, Australia

Chris Greening

14. ARC Research Hub for Carbon Utilisation and Recycling, Monash University, Clayton, Victoria, Australia

Chris Greening

## Contributions

C.G., R.G., A.K. and G.M.C. conceived and designed the project. C.G. and R.G. administered the project. A.K. and R.G. generated *M. smegmatis* mutants, purified proteins, performed kinetic analysis and $H_2$ consumption assays. H.V., R.G., A.K. and M.B. prepared cryo-EM grids and collected and processed cryo-EM data. R.G. modelled and analysed the Huc structures. G.B., K.A.V., P.R.C. and Z.D. devised, performed and analysed

the PFE experiments. G.B., P.H. and H.S.S. devised, performed and analysed the EPR experiments. G.B., M.S. and S.T.S. devised, performed and analysed the FTIR experiments. S.K., J.B., R.J. and R.G. designed, performed and analysed the molecular dynamics simulations. R.B.S. and C.K.B. designed and performed the mass spectrometry experiments. R.G., C.G. and A.K. wrote and edited the manuscript with input from all authors.

### Corresponding authors

Correspondence to Rhys Grinter or Chris Greening.

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Peer review

### Peer review information

*Nature* thanks Martin Högbom, Janet Vonck and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

### Extended Data Fig. 1 Huc isolation and purification.

(a) Left panel: A chromatogram showing streptactin-purified Huc separated via size exclusion chromatography on a Superose 6 10/300 column. The green highlighted region contains the Huc oligomer, while the pink region contains a low molecular weight Huc species. Right panel: A Coomassie-stained SDS-PAGE gel showing fractions from the coloured peak regions of the chromatogram. (b) A schematic of the Huc gene cluster showing the location of *hucM* (MSMEG_2261) compared to *hucS* and *hucL*. (c) A native-PAGE gel of the purified Huc oligomer (green) and low molecular species (pink), stained with Coomassie (left panel) and NBT (right panel). (d) Differential scanning fluorimetry analysis of the Huc oligomer showing a transition in tryptophan-derived fluorescence at 78.3 °C, indicating the melting temperature of the oligomer. (e) An NBT-stained native-PAGE gel of *M. smegmatis* cell lysate showing that Huc does not form an oligomer in the *ΔhucM* strain. (f) Gas chromatography analysis of the $H_2$ concentration of the headspace of sealed vials containing Huc, *E. coli* Hyd1 or no enzyme, showing Huc can oxidize $H_2$ from 100 ppm to the limit of detection for the gas chromatograph (40 ppbv) (green line). In contrast, Hyd1 does not oxidise $H_2 < 8$ ppmv. Data are presented as mean values +/− SD.

Source data

## Extended Data Fig. 2 Comparison of protein film electrochemistry response of Huc with Hyd1 and Hyd2 from *E. coli* at different $H_2$ concentrations.

(a) The Huc onset potential calculated from data presented in Fig. 1f, and the calculated ${{\rm{E}}^{{\rm{0}},}}_{{\rm{2}}{\rm{H}}^{+}/{\rm{H}}_{{\rm{2}}}}$ potential as a function of $H_2$ concentration, showing that the difference between these potentials (the overpotential, η) remains constant regardless of $H_2$ concentration. Data are presented as mean values +/− SD (b) The behavior of *E. coli* Hyd2 and Hyd1 at low $H_2$. Cyclic voltammograms were recorded with immobilized hydrogenase with a gas mixture flushing through the headspace comprising argon, $N_2$ and the indicated level of $H_2$. Top: *E. coli* Hyd2; Bottom: *E. coli*

Hyd1. The dashed vertical line indicates the potential of the $H^+/H_2$ couple, calculated for pH 7, 5% $H_2$ at ambient pressure. (c) Cyclic voltammograms with immobilized hydrogenase, in enzyme-free mixed buffer (pH 7.0, 25 °C) with a gas mixture flushing through the headspace comprising argon, $N_2$ and the indicated level of $H_2$. The dashed vertical line indicates the potential of the $H^+/H_2$ couple, calculated for pH 7, 5% $H_2$ at ambient pressure. For each enzyme, the set of scans was recorded on the same enzyme film in the order of low to high $H_2$. Top panel: Huc; middle panel: *E. coli* Hyd1; bottom panel: *E. coli* Hyd2. At 5% $H_2$, Hyd1 shows an onset potential of −340 mV for $H_2$ oxidation, more negative than the onset observed for Huc of −120 mV. In contrast, Hyd2 shows reversible catalysis at 5% $H_2$ with the current for $H^+$ reduction and $H_2$ oxidation cutting the zero current potential very close to the thermodynamic ${{\rm{E}}}_{{{\rm{2}}{\rm{H}}}^{+}/{{\rm{H}}}_{{\rm{2}}}}$ potential of −375 mV. The $H^+$ reduction current for Hyd2 increases in magnitude as the $H_2$ concentration is lowered and product inhibition is relieved.

Source data

## Extended Data Fig. 3 Cryo-EM visualisation and 3D reconstruction of the Huc oligomer.

(a) Motion-corrected micrographs of vitrified purified Huc oligomers, showing Huc associated with membrane vesicles (left panel) and free Huc (right panel). (b) Selected 2D class averages of Huc oligomer, showing C4 symmetry and flexible membrane-associated stalk. (c) Data processing workflow for the Huc oligomer reconstruction. (d) The Euler angle distribution of particles used for Huc oligomer reconstruction. (e) Gold-standard Fourier shell correlation (FSC) curves calculated from two independently refined half-maps indicate an overall resolution of 2.19 Å at FSC = 0.143, and Guinier plot indicates a sharpening B-factor of 57.2. (f) A local resolution map of the Huc oligomer, showing a resolution range of ~1.8 to 4.8 Å from the Huc core to the periphery, indicative of significant interdomain motion.

## Extended Data Fig. 4 Map quality of Huc oligomer and dimer Cryo-EM reconstructions and Cryo-EM density map and AlphaFold model of the Huc stalk region.

(a) Maps from the 2.19 Å Huc oligomer reconstruction for one Huc lobe ($HucS_2L_2$) and 2 HucM molecules. (b) Helices from HucS and HucM with associated density from oligomer reconstruction. (c) Maps from the 1.52 Å Huc dimer reconstruction. (d) Examples of density corresponding to amino acids from the Huc dimer reconstruction contoured at 5 or 15 σ. (e) The stalk region from Cryo-EM density maps from Supplementary Fig. 4. (f) A composite of the Huc stalk and body map regions from the 2.19 Å dataset. (g) A cut-away view of the composite map from panel f, showing the enclosed chamber and the centre of the Huc complex. (h) A surface view of the AlphaFold model of the HucM C-terminal stalk region shown as a cutaway electrostatic surface (left) and the internal surface lined with hydrophobic residues (right). (i) The HucM C-terminal region as in panel d, showing the positively charged region at the base of the stalk (left) and the high frequency of arginine and lysine residues in this region (right).

## Extended Data Fig. 5 Assignment of HucS iron sulfur clusters as 3×[3Fe-4S] and the Ni-B state of oxidized Huc.

(a) Cross-eye stereo images of carved density maps corresponding to the three HucS [FeS]-clusters (yellow) and the additional non-cystine [FeS]-cluster interacting residues (red). Density maps are derived from the 1.52 Å Huc dimer reconstruction contoured to 5 σ. (b) EPR spectra collected from a Huc solution in the as-isolated oxidized state (Huc-Air) and $H_2$ reduced state (Huc-$H_2$) at low T (7 K); Modulation frequency 100 kHz, amplitude 10 G; Microwave frequency 9.4 GHz; power 2 mW. Spectra collected at low T of $H_2$-flushed samples showed no discernible contribution from rhombic components typically associated with reduced $[4Fe\text{-}4S]^+$ clusters. (c) Zoomed-in view of the EPR spectra of as-isolated Huc measured at 10 K (black) and 40 K (green). The presence of multiple peaks at low T is consistent with strong spin-spin interactions between the EPR-active Ni center and the $[3Fe\text{-}4S]^+$ cluster, giving rise to the “split” nickel signals

observed. At higher temperatures, these split signals collapse, and a spectrum more characteristic of a canonical Ni-B signal is recovered. The simulation of a single species with g = 2.36, 2.22, and 2.00 is overlaid; we note that the third g-value is unresolved because of overlap with the cluster signal. Additional split features of the $[3Fe\text{-}4S]^+$ cluster are indicated at lower g-values. (d) Zoomed-in view comparing the EPR spectra of as-isolated (black) and $H_2$-reduced (red) Huc at 7 K. The nickel signals are seen to shift considerably upon $H_2$ incubation. The observed reduction in spectral anisotropy is consistent with the changes expected in converting from a Ni-B to a Ni-C species, though some splitting of the signals are still observed.

Source data

## Extended Data Fig. 6 Molecular dynamics simulations of HucSL in the presence of $H_2$ or $O_2$.

(a) A zoomed view of the large and small subunits of an [NiFe]-hydrogenase from *Desulfovibrio vulgaris* (PDB ID = 4U9H)[24], showing the location and width of the hydrophobic gas channels that provide substrate access to the [NiFe] active site. See Supplemental Table 2 for tunnel width statistics. (b) Cumulative occupancy plots showing the proximity of $H_2$ and $O_2$ molecules to the [NiFe] cluster of Huc throughout the molecular dynamics simulations. (c) An expanded view of the positions of a representative subset of $O_2$ molecules in closest proximity to the Huc [NiFe] cluster in molecular dynamics simulations, with the path of the hydrophobic gas channels shown as blue, green, and cyan spheres. (d) A zoomed stereo view of $O_2$ as described in panel c. (e) and (f) The positions of a representative subset of $H_2$ molecules in closest proximity to the Huc [NiFe] cluster are displayed as described in panels c and d. (g) A distance-over-time plot showing the relative proximity of serine 104 to arginine 443, in the Huc MD simulations in the presence of $H_2$. (h) A distance-over-time plot showing the relative proximity of serine 104 to arginine 443, in the Huc simulations in the presence of $O_2$. (i) The relative position of arginine 443

in states 1 (top panel) and 2 (bottom panel) populated during the Huc MD simulations.

## Extended Data Fig. 7 Molecular dynamics simulations showing oxygen accessibility to the catalytic site of Huc gas channel mutants.

(a) Distance-over-time plots showing the relative proximity of $H_2$ and $O_2$ to the [NiFe] cluster of WT and gas channel mutant Huc variants in the Huc MD simulations. (b) Cumulative occupancy plots showing the proximity of $H_2$ and $O_2$ molecules to the [NiFe]-cluster of WT and gas channel mutant Huc variants throughout the molecular dynamics simulations. (c) A structural representation of the region of Huc encompassing the catalytic cluster from the WT and gas channel mutant Huc variants from Run 1 of the MD simulations, showing the position of $O_2$ molecules in this region at each frame of the simulations. Residues targeted for mutagenesis are shown as sticks, and regions of enhanced $O_2$ access in the E15A + I64A and E15A + I64A + L112A mutants are indicated.

## Extended Data Fig. 8 FTIR analysis of Huc in the presence of ambient air, $N_2$ or $H_2$.

(a) Absolute FTIR spectra of Huc equilibrated in air, followed by sequential equilibration in a 100% atmosphere of the labelled gas, followed by air. [NiFe] cluster states are assigned based on literature-derived values (see Supplementary Table 3). (b) Difference spectra derived from spectra shown in panel a. (c) A time-course plot showing the relative fractions of the FTIR assigned states immediately following the transfer of Huc from ambient air to a 100% $H_2$ atmosphere, the catalytic Ni-R state becomes populated at the expense of hydroxide bound Ni-B over a time scale of minutes. (d) A time-course plot showing the relative fractions of the FTIR assigned states immediately following the transfer of Huc in a mixed Ni-SI, Ni-R state from 100% $N_2$ to a 100% $H_2$ atmosphere. The catalytic Ni-R state becomes populated at the expense of the non-hydrogen bound reduced Ni-SI state over a timescale of seconds. (e) A time-course plot of [NiFe] cluster states

following the transfer of Huc predominantly in the Ni-SI state from 100% $N_2$ into 80%:20% $N_2$ to $O_2$. Initially, the Ni-SI state is populated at the expense of the hydrogen-bound Ni-C and Ni-R states, followed by population of the Ni-B state at the expense of Ni-SI over a timescale of minutes. (f) A time-course plot of [NiFe] cluster states of Hyd2 from *E. coli* following the transfer from 100% $N_2$ into 99%:1% $N_2$ to $O_2$. The Ni-B state is rapidly populated at the expense of all other states over a timescale of seconds. Data from Senger et al. 2018[85]. (g) A scheme of the proposed catalytic cycle for [NiFe]-hydrogenases, showing the states occupied by Huc during FTIR analysis. (h) Vibrational frequencies for the CO and CN bands identified from FTIR spectra for each of the proposed Huc states. Peak position and FWHM units = $cm^{-1}$.

Source data

## Extended Data Fig. 9 Comparison of Huc with an $O_2$ sensitive [NiFe]-hydrogenase.

(a) A stereoview of the Huc [NiFe] cluster and surrounding amino acids. Amino acids conserved in $O_2$-sensitive hydrogenases and shown in blue, while divergent residues are shown in red. Amino acid type and position are labelled for divergent residues. (b) A stereoview overlay of the Huc catalytic site residues from panel a with those of the $O_2$-sensitive hydrogenase from *D. vulgaris* (PDB ID = 4U9H). Huc residues are coloured as in panel a, while *D. vulgaris* hydrogenase residues are coloured cyan (conserved) and pink (divergent). (c) A sequence alignment of Huc and *D. vulgaris* hydrogenase catalytic site residues, with divergent residues highlighted in red. (d) A cross-eye stereo view of D-histidine 166 from HucL shown as sticks demonstrating amino acid chirality. The corresponding Cryo-EM density map is shown at 5 σ, demonstrating the density is consistent with the D-isomer (top panel). L-histidine 194 form HucL is shown for reference (bottom panel). (e) The location of L-histidine 235 from the *D. vulgaris* hydrogenase, which corresponds to D-histidine 166 from Huc, showing its interaction with the proximal 4Fe-4S cluster. (f) An overlay of the relative positions of D-histidine 166 from Huc and L-

histidine 235 from the *D. vulgaris* hydrogenase showing how the D-isomer shifts the position of the main chain in Huc.

**Extended Data Table 1 Cryo-EM data collection, refinement and validation statistics**

# Supplementary information

## Supplementary Information

This file contains Supplementary Figs. 1–5; Supplementary Tables 1–4; Supplementary Notes 1–4 and references.

## Reporting Summary

## Peer Review File

## Supplementary Data (Supplementary Figure 4)

This file contains source data for Supplementary Fig. 4.

## Supplementary Video 1

3D variability analysis of Huc reveals significant inter-lobe movement.

## Supplementary Video 2

Coordination and density map for the Huc proximal [3Fe-4] cluster.

## Supplementary Video 3

Coordination and density map for the Huc medial [3Fe-4] cluster.

## Supplementary Video 4

Coordination and density map for the Huc distal [3Fe-4] cluster.

**Supplementary Video 5**

Reconstruction of the structure of the full-length Huc complex.

**Supplementary Video 6**

Density map for the region of the HucL polypeptide containing d-His166

**Supplementary Video 7**

Density map for the region of the HucL polypeptide containing l-His193.

**Supplementary Video 8**

The electron transport relay of Huc

**Supplementary Video 9**

The Huc menaquinone binding site is gated by Tyr229 of HucS.

**Supplementary Video 10**

Huc possesses an internal hydrophobic chamber for reducing menaquinone.

## Source data

**Source Data Fig. 1**

**Source Data Fig. 3**

**Source Data Fig. 4**

**Source Data Extended Data Fig. 1**

**Source Data Extended Data Fig. 2**

## Source Data Extended Data Fig. 5

## Source Data Extended Data Fig. 8

# Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

# About this article

## Cite this article

Grinter, R., Kropp, A., Venugopal, H. *et al.* Structural basis for bacterial energy extraction from atmospheric hydrogen. *Nature* **615**, 541–547 (2023). https://doi.org/10.1038/s41586-023-05781-7

- Received: 17 May 2022
- Accepted: 02 February 2023
- Published: 08 March 2023
- Issue Date: 16 March 2023

- DOI: https://doi.org/10.1038/s41586-023-05781-7

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05781-7

- Matters Arising
- Open Access
- Published: 15 March 2023

# Common orthopaedic trauma may explain 31,000-year-old remains


- Nicholas J. Murphy,
- Joshua S. Davis,
- Seth M. Tarrant &
- …
- Zsolt J. Balogh

Show authors


*Nature* **volume 615**, pages E13–E14 (2023)


- 1497 Accesses
- 14 Altmetric
- Metrics details


Matters Arising to this article was published on 15 March 2023

The Original Article was published on 07 September 2022

arising from T. R. Maloney et al. *Nature* https://doi.org/10.1038/s41586-022-05160-8 (2022)

The fascinating discovery of skeletal remains in Borneo of an individual (TB1) with absent left distal tibia, fibula and foot from 31,000 years ago[1] has been proposed as evidence of a contemporaneous sophisticated amputation procedure. Maloney et al.[1] infer from the bony abnormalities that surgical amputation is the only possible explanation and, furthermore, that the limb shows no evidence of infection. We dispute the conclusion that these skeletal remains provide evidence of a transosseous surgical amputation and that the limb shows no signs of infection. We propose that the skeletal findings have more plausible alternative explanations, such as the natural history of an injury pattern commonly encountered in blunt orthopaedic trauma, an open distal tibia/fibula fracture with growth-plate involvement.

From the perspective of orthopaedic trauma surgeons practicing in this field, the exclusion of blunt trauma as a potential mechanism of injury is a rather reductionist approach to a differential diagnostic puzzle with several missing pieces from thousands of years ago. Maloney et al.[1] dismissed blunt trauma with the assertion that it "typically causes comminuted and crushing fractures". To support this statement, they cited an isolated case report of an axis fracture[2] (the second cervical vertebra in the neck). We disagree with the statement and suggest the supporting citation is inadequate. Oblique fracture patterns such as TB1's are frequently observed from blunt trauma in clinical practice[3].

Physeal (growth plate) fractures of the distal tibia and fibula are common injuries in adolescents[4]. The most common subtype[5] involves fracture through the distal tibial physis and into the metaphysis (Salter–Harris type II[6]), and typically occurs concomitantly with a distal fibula fracture[7]. It is not uncommon that the medial apex of the fracture pierces through the skin[8,9], while the foot displaces laterally, as seen in Fig. 1. The mean age of people presenting with this Salter–Harris type II injury to the distal tibia and fibula is 12–13 years[7], which corresponds to the predicted age of injury for the individual of whom the skeletal remains were discovered in Borneo, given he or she is estimated to have died aged 19–20 (a predicted 6–9 years after suffering the injury)[1]. The typical mechanism of these injuries is forced inversion or eversion while the foot is fixed in position on the

ground. In Borneo 31,000 years ago, this may have eventuated from any slip or misstep whilst running, or a jump from low height. Today, open ankle fractures are managed with intravenous antibiotics, tetanus prophylaxis and expeditious debridement with open reduction and internal fixation in the operating theatre.

**Fig. 1: Plain radiograph of a distal tibia and fibula fracture.**



A plain radiograph of a distal tibia and fibula fracture involving the physis (growth plate) in a 14-year-old individual.

In Borneo 31,000 years ago, the natural history of an open physeal ankle fracture without modern surgical care could quite plausibly have produced the findings encountered in TB1's skeletal remains. In some cases, the inoculation of bacteria into exposed bone could result in acute infection progressing to overwhelming sepsis and death. In other cases, a chronic osteomyelitis can develop, often with the formation of a life-long draining sinus[10]. Survival with chronic osteomyelitis was described on many occasions in the pre-antibiotic era[11]. In the 1830s, Nathan Smith, a professor of surgery at Yale University, suggested that most people with osteomyelitis he observed survived with the condition, writing "a very great majority of patients survive the attack, albeit with long confinement, protracted suffering and great emaciation."[12]. The "long confinement, protracted suffering" was most probably TB1's fate during his young adulthood. It is highly implausible that either surgical amputation or an open fracture, in the absence of antiseptics, anaesthetics or antibiotics, could occur without a subsequent established infection. Furthermore, the bony changes shown in figure 3 of Maloney et al.[1] are typical of chronic osteomyelitis—the cortical thickening of the distal tibia and fibula are consistent with involucrum, and the small bony defect of the distal tibia could represent an area of sequestrum. The evidence of bone lysis and necrosis at the distal tibial and fibula, which Maloney et al.[1] refer to in the caption of figure 3b, could quite plausibly occur secondary to infection[13]. The suggested mechanism explains other characteristics of the skeletal findings—the missing malleolar parts of the distal tibia and fibula are consistent with common adolescent fracture patterns, and the small size of the left tibia and fibula relative to the right is highly suggestive of physeal arrest, which is a common complication of displaced physeal fractures[7,14].

With a distal tibia and fibula fracture already present and necrosis of the bone and surrounding soft tissues occurring due to infection, terminalization of the limb—that is, cutting through the remaining soft tissues—is a more plausible scenario. This is a substantially different proposition from the primary transosseous surgical amputation described in Maloney et al.[1], which states that the bone must have been cut with a sharp

instrument. It is impossible to know whether loss of the foot occurred around the time of injury, or weeks to months later. If an arterial injury accompanied the initial bony injury, and the limb suffered distal ischaemia as a result, a dry gangrenous process may even have autoamputated the limb without any assistance.

We cannot exclude the possibility of rarer causes to explain these skeletal remains—for example, congenital transverse deficiency of the lower limb, a rare congenital anomaly that can manifest as a hypoplastic limb with absent foot[15]. If TB1 were born with this condition, weightbearing on the footless lower limb without a durable heel pad could have caused ulceration and the chronic infective changes that we observe in the bony architecture of TB1's distal tibia and fibula.

Overall, we find that the conclusions drawn by Maloney et al.[1] are unconvincing. Performing primary supra-articular transosseous surgical amputation through the thick cortices of the tibia without specialized metallic tools (at least chisel and saw) would be very difficult and is highly improbable. If the people in Borneo were performing lower-limb amputation using 'sharp instruments', it would have been easier to perform transarticular amputation through the soft tissues of the ankle joint, where it is not necessary to transect thick cortical bone. This is not the pattern observed in these skeletal remains.

We suggest that interdisciplinary input from expert orthopaedic trauma surgeons and bone and joint infection experts would be of value in archaeological studies such as this to aid in formulating plausible explanations of injury mechanism and infectious processes, as palaeopathology is unlikely to cover the breadth of specialized understanding required. Unfortunately, our concerns are not limited to the explanation of the missing part of the skeleton. Figure 3a of Maloney et al.[1] contains a photographed reconstruction of TB1's bony anatomy, with the distal portion of the right tibia placed back-to-front such that the medial malleolus is incorrectly articulating with the distal fibula. As an image that is likely to be frequently reproduced, and that has already had considerable media attention, this requires correction. Although we cannot support the conclusions of Maloney et al.[1], we nevertheless consider the findings

described to be of great interest. Even if TB1 did not undergo a transosseous surgical amputation, the findings demonstrate evidence of an individual who, 31,000 years ago, must have had enormous kin support to survive for several years after a severe open injury, which of itself seems a notable detail about our ancestors.

## Reporting summary

Further information on experimental design is available in the Nature Portfolio Reporting Summary linked to this Article.

## Data availability

All relevant data are included in the Article.

## References


1. Maloney, T. R. et al. Surgical amputation of a limb 31,000 years ago in Borneo. *Nature* **609**, 547–551 (2022).

2. Aydin, K. & Cokluk, C. A fracture of unilateral pars interarticularis of the axis: a case report. *Turk. Neurosurg.* **17**, 155–157 (2007).

3. Domzalski, M. E., Lipton, G. E., Lee, D. & Guille, J. T. Fractures of the distal tibial metaphysis in children: patterns of injury and results of treatment. *J. Pediatr. Orthop.* **26**, 171–176 (2006).

4. Podeszwa, D. A. & Mubarak, S. J. Physeal fractures of the distal tibia and fibula (Salter-Harris type I, II, III, and IV fractures). *J. Pediatr. Orthop.* **32**, S62–S68 (2012).

5. Spiegel, P. G., Cooperman, D. R. & Laros, G. S. Epiphyseal fractures of the distal ends of the tibia and fibula. A retrospective study of two hundred and thirty-seven cases in children. *J. Bone Joint Surg. Am.* **60**, 1046–1050 (1978).

6. Salter, R. B. & Harris, W. R. Injuries involving the epiphyseal plate. *J. Bone. Joint Surg.* **45**, 587–622 (1963).

7. Russo, F., Moor, M. A., Mubarak, S. J. & Pennock, A. T. Salter-Harris II fractures of the distal tibia: does surgical management reduce the risk of premature physeal closure? *J. Pediatr. Orthop.* **33**, 524–529 (2013).

8. Nandra, R. S., Wu, F., Gaffey, A. & Bache, C. E. The management of open tibial fractures in children. *Bone Joint J.* **99-B**, 544–553 (2017).

9. Simske, N. M., Audet, M. A., Kim, C.-Y. & Vallier, H. A. Open ankle fractures are associated with complications and reoperations. *OTA Int.* **2**, e042 (2019).

10. Panteli, M. & Giannoudis, P. V. Chronic osteomyelitis: what the surgeon needs to know. *EFORT Open Rev.* **1**, 128–135 (2016).

11. Klenerman, L. A history of osteomyelitis from the *Journal of Bone and Joint Surgery*: 1948 to 2006. *J. Bone Joint Surg. Br.* **89**, 667–670 (2007).

12. Smith, N. *Medical and Surgical Memoirs* 97 (WA Francis, 1831).

13. Ciampolini, J. & Harding, K. G. Pathophysiology of chronic bacterial osteomyelitis. Why do antibiotics fail so often?. *Postgrad. Med. J.* **76**, 479–483 (2000).

14. Herman, M. J. & Dean MacEwen, G. Physeal fractures of the distal tibia and fibula. *Curr. Orthop.* **17**, 56–62 (2003).

15. Gold, N. B., Westgate, M.-N. & Holmes, L. B. Anatomic and etiological classification of congenital limb deficiencies. *Am. J. Med. Genet. A* **155**, 1225–1235 (2011).

# Author information

## Authors and Affiliations

1. Department of Traumatology, John Hunter Hospital & University of Newcastle, Newcastle, New South Wales, Australia

   Nicholas J. Murphy, Seth M. Tarrant & Zsolt J. Balogh

2. School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia

   Joshua S. Davis

3. Department of Infectious Diseases, John Hunter Hospital, Newcastle, New South Wales, Australia

   Joshua S. Davis

### Contributions

Z.J.B. initiated this Comment. N.J.M. prepared the first draft manuscript. All of the authors revised and approved the final manuscript.

### Corresponding author

Correspondence to Zsolt J. Balogh.

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Supplementary information

Reporting Summary

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

## About this article

### Cite this article

Murphy, N.J., Davis, J.S., Tarrant, S.M. *et al.* Common orthopaedic trauma may explain 31,000-year-old remains. *Nature* **615**, E13–E14 (2023). https://doi.org/10.1038/s41586-023-05756-8

- Received: 24 September 2022
- Accepted: 25 January 2023
- Published: 15 March 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05756-8

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05756-8

- Matters Arising
- Open Access
- Published: 15 March 2023

# Reply to: Common orthopaedic trauma may explain 31,000-year-old remains


- Melandri Vlok ,
- Tim Maloney ,
- India Ella Dilkes-Hall,
- Adhi Agus Oktaviana,
- Pindi Setiawan,
- Andika Arief Drajat Priyatno,
- Marlon Ririmasse,
- I. Made Geria,
- Muslimin A. R. Effendy,
- Budy Istiawan,
- Falentinus Triwijaya Atmoko,
- Shinatria Adhityatama,
- Ian Moffat,
- Renaud Joannes-Boyau,
- Adam Brumm &
- …
- Maxime Aubert

Show authors


*Nature* **volume 615**, pages E15–E18 (2023)

- 885 Accesses

- 7 Altmetric
- Metrics details

The Original Article was published on 15 March 2023

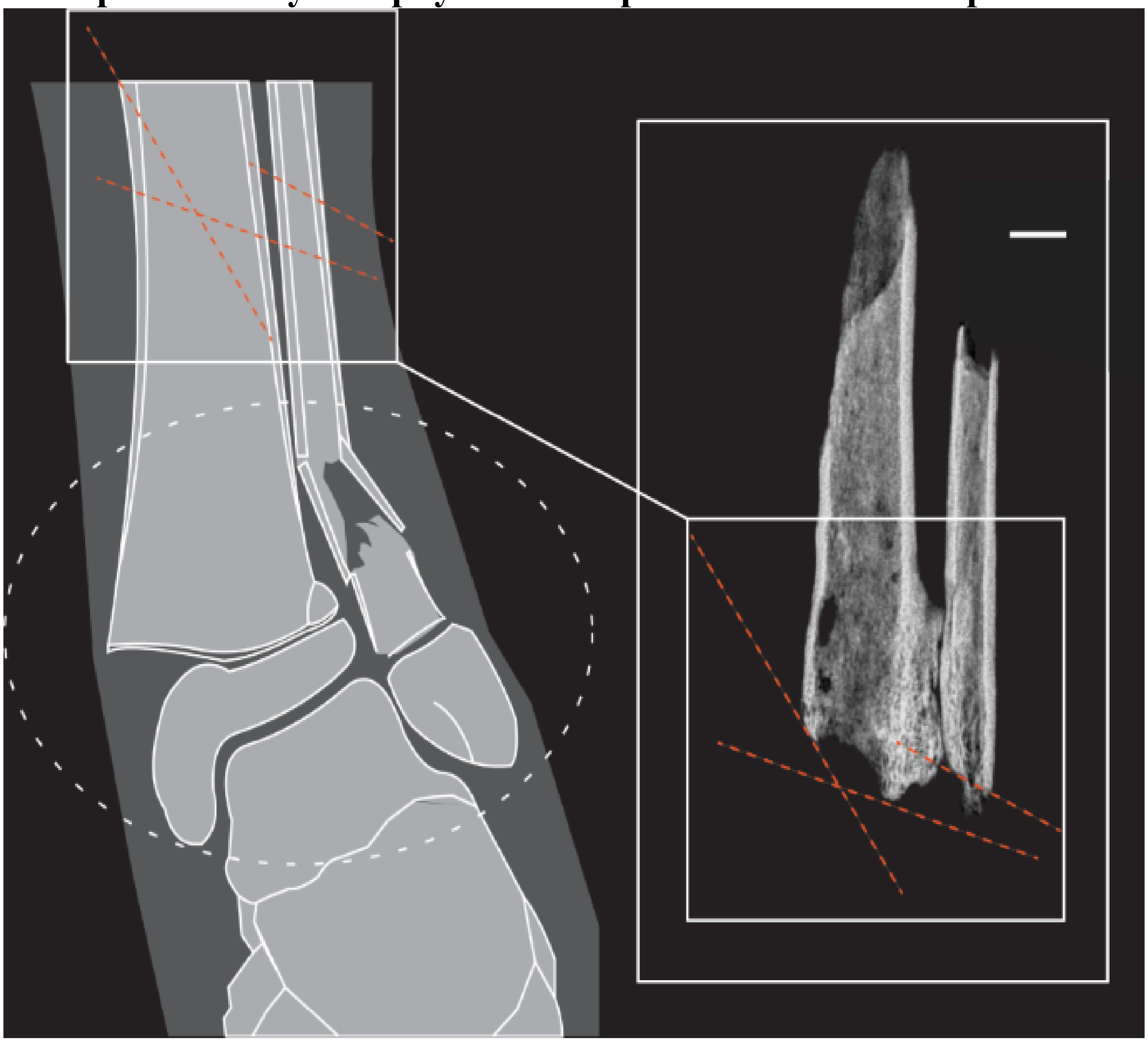
reply to Murphy et al. *Nature* https://doi.org/10.1038/s41586-023-05756-8 (2023)

We appreciate the accompanying technical Comment by Murphy et al.[1]—a group of practicing orthopaedic surgeons—on our original paper[2]. However, we strongly disagree with their conclusion that a reductionist approach was taken in the diagnosis of surgical amputation in a 31,000-year-old individual (TB1) from Borneo. We note that a complete systematic differential diagnosis was indeed completed (Extended Data Table 1); this process involved careful consideration of the most common and banal conditions first, such as accidental fracture, before considering the possibility of more rare and unusual circumstances. Through this iterative process, fracture was first eliminated as a possibility, followed by natural causes of amputation.

Surgical amputation was the remaining scenario left that completely described the characteristics that we observed in the bone. As is standard for palaeopathological analysis, a detailed description of the pathology was undertaken, including recording of the location and aspect of affected bone, the type of bone affected, the mechanism of injury, the degree of healing, complications to healing, force and fracture type. This detailed analysis means that certain aspects of trauma were excluded from the differential diagnosis due to the specific location of the injury. It was at this stage that physeal fractures, which Murphy et al.[1] correctly recognize as common fractures in early adolescence, were disregarded from the differential diagnosis as the affected portion of the bone was at the site of the mid to distal lower third diaphysis and not near to the distal metaphyseal region (Fig. 1). We surmise that Murphy et al.[1] may have mistaken the thin cortices of TB1's tibia and fibula for that at the diaphyseal-to-metaphyseal transition in the bone that is naturally thin. However, with TB1, the cortices in these bones (and indeed also the left femur) are thin due to extreme atrophy that

probably occurred over a number of years. We acknowledge that two-dimensional photographs and radiographs can misrepresent to readers injuries that, in reality, occur in three dimensions. Thus, we provide publicly available three-dimensional computed tomography files of the amputation.

**Fig. 1: A Salter–Harris type II fracture of left tibia and fibula similar to the one presented by Murphy et al. compared with TB1's amputation.**



The amputation site is more proximal (white box) to the region of Salter–Harris physeal fractures (white dashed oval). At minimum, three different angles of force are present in TB1's amputation (red dashed lines) as opposed to one angle of force in Salter–Harris type II. Medial physis fracture of the tibia is absent in TB1 as is necessary for Salter–Harris type II

classification with fibular involvement (see figure 1 of Murphy et al.[1]). Scale bar, 10.0 mm.

Moreover, the age of 6 to 9 years after surgery is a minimum age based on the minimal timing required for the completion of bone remodelling in the major long bones and, given the size of the lower limb bones, it is probable that the injury occurred in childhood. As Murphy et al.[1] are aware, physeal stasis can have diverse traumatic origins as well as stasis of longitudinal growth in general[3,4]. Experimental animal studies demonstrate the importance for muscular activity to initiate longitudinal growth of bones through biomechanical strain. Thus, the small size of the left limb bones can be readily attributed to the existing evidence for bone atrophy related to muscle wastage[5].

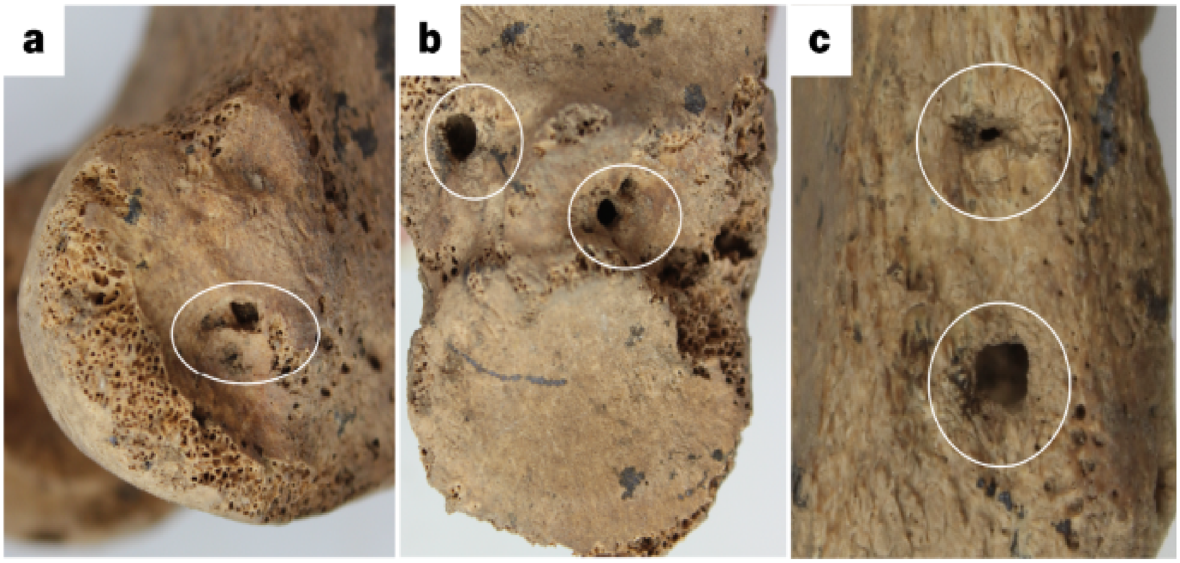
We are uncertain what Murphy et al.[1] are referring to in their second paragraph when relating the cervical fracture to the force applied to the amputation site. Although it is possible that the cervical vertebral fracture occurred in the same event that led to the need for amputation of the lower limb, the limitations of bone response prevent us from investigating this possibility any further. Owing to the lack of empirical evidence, we refrain from speculating on the motivation or underlying cause that led to the decision to amputate. It is of course possible that the trauma described by Murphy et al.[1] was the ultimate mechanism of injury that led to the child's limb being surgically amputated at the location of the distal diaphysis. We clarify that we are not saying oblique fractures of the long bone shafts do not occur from blunt force trauma but are atypical in cases from an accident (excluding modern situations including transport), particularly one where the fibula and tibia were both fractured.

Murphy et al.[1] point out in detail the requirements for their proposed scenario to have occurred but do not see the improbability of such a condition in the context of the Pleistocene tropics of Borneo. They do suggest soft tissue-only surgery as an alternative that would have involved antisepsis and debridement, which is arguably a far more sophisticated (and therefore less parsimonious) form of care that would have required a complex understanding of the anatomical basis for infection to specifically remove the infected tissue (rather than performing an entire amputation). If

the fracture was not reduced through fixation, as is the case in modern Western surgical practices, a dead foot would have probably been an extreme impediment for the rugged mountainous terrain, and far more painful than a stump. Moreover, the fractured foot would have been susceptible to repeated infection as it was carried throughout the environment.

Murphy et al.[1] incorrectly describe the remodelled bone as osteomyelitis. To support their argument, they report an anecdote in a review on the history of osteomyelitis that is from a single memoir of an American surgeon published in 1831, whose patients received treatment in hospital[6]. Osteomyelitis in the tropics is more aggressive owing to the greater diversity of the pathogens that cause osteomyelitis, the suitability of *Staphylococcus aureus*—the most common cause of osteomyelitis—to the humidity in the tropical belt and, potentially, the reduced amount of clothes worn in tropical environments increasing the infection risk of exposed wounds[7,8]. Although the mortality rate of untreated sepsis is not documented in the tropics, antibiotic-era in-hospital mortality rates in post-amputation contexts are reported to be as high as 10% and, in the Vietnam War, sepsis was attributed to 12% of deaths in surgical patients, the third leading cause of mortality in that conflict[9,10,11]. Osteomyelitis, both pyogenic and non-pyogenic, is readily observed in archaeological bone. In pyogenic forms, death of bone leading to sequestrum is readily observable surrounded by a shell of bone known as involucrum. Cloacae—pus draining holes—form to drain the pus from the medullary canal. Although there are circular holes in the bone, these are clearly a result of carnivore puncture and beetle scavenging marks, which are very common causes of post-mortem skeletal damage observed in Southeast Asian archaeological human skeletons (Fig. 2).

**Fig. 2: Cloacae of the right femur and tibia compared to carnivore puncture holes of the left tibia.**




**a**–**c**, Comparison of cloacae of the right femur (**a**) and tibia (**b**) with the carnivore puncture holes of the left tibia (**c**). The left tibia holes are clearly caused by punctures in dry bone resulting in square jagged margins to the cavities. By contrast, the margins of the holes in the right tibia and femur are rounded due to the constant remodelling process in the development of the cloacae. The femoral cloaca (**a**) also presents with a clear lytic channel consistent with infection.

The radiographs of TB1's amputated limb (Fig. 1) show a lack of bone radiolucency associated with the development of sequestra, and the localization of radiodense bone only intermediate to the tibia and fibula is consistent with myositis ossification, and not with osteomyelitis, which will result in subperiosteal inflammation and subsequent new bone development on a more diffuse scale around the infected site. The complete lack of subperiosteal change to the tibia and fibula away from the ossified region, as well as the initiation of the subperiosteal new bone, from both the tibia and the fibula, to meet intermediately, is consequently not consistent with osteomyelitis. Moreover, chronic osteomyelitis is expected to be associated with some level of continued subperiosteal activity observed as woven bone and, in this case, the bone is entirely lamellar. Evidence of osteomyelitis in the right limb is available for comparison as well as dry bone examples from prehistoric Southeast Asia associated with and without fracture[12,13].

We do concede the error that the medial malleolus of the right tibia is not placed in anatomical position in figure 3a of our original paper[2]. However, the aim of this figure is to represent the general completeness of the skeleton, and the relationship of the size of the left and right limbs, which we believe the figure succeeds in presenting. Given the medial malleolus is barely discernible, we believe the matter of anatomical correctness to be negligible.

## Reporting summary

Further information on experimental design is available in the Nature Portfolio Reporting Summary linked to this Article.

## Data availability

CT data are available at Figshare (https://figshare.com/projects/CT_Data_Tebo_TB1_Borneo_Kalimantan/150765).

## References


1. Murphy, J., Davis, J., Tarrant, S. & Balogh, Z. Common orthopaedic trauma may explain 31,000-year-old remains. *Nature* https://doi.org/10.1038/s41586-023-05756-8 (2023).

2. Maloney, T. R. et al. Surgical amputation of a limb 31,000 years ago in Borneo. *Nature* **609**, 547–551 (2022).

3. Westberry, D. E., Davids, J. R. & Pugh, L. I. The Boyd amputation in children: indications and outcomes. *J. Pediatr. Orthop.* **34**, 86–91 (2014).

4. Krajbich, I. J. Lower-limb deficiencies and amputations in children. *JAAOS* **6**, 358–367 (1998).

5. Hall, B. K. & Herring, S. Paralysis and growth of the musculoskeletal system in the embryonic chick. *J. Morphol.* **206**, 45–56 (1990).

6. Smith, N. *Medical and Surgical Memoirs* (WA Francis, 1831).

7. Schaumburg, F., Alabi, A., Peters, G. & Becker, K. New epidemiology of *Staphylococcus aureus* infection in Africa. *Clin. Microbiol. Infect.* **20**, 589–596 (2014).

8. Munshi, B., MacFater, W., Hill, A. G. & McCaig, E. H. Paediatric osteomyelitis in Fiji. *World J. Surg.* **42**, 4118–4122 (2018).

9. Bickler, S. W. & Sanno-Duanda, B. Bone setter's gangrene. *J. Pediatr. Surg.* **35**, 1431–1433 (2000).

10. Shittu, A. et al. Tropical pyomyositis: an update. *Trop. Med. Int. Health* **25**, 660–665 (2020).

11. Blyth, D. M., Yun, H. C., Tribble, D. R. & Murray, C. K. Lessons of war: combat-related injury infections during the Vietnam War and Operation Iraqi and Enduring Freedom. *J. Trauma Acute Care Surg.* **79**, S227–S235 (2015).

12. Scott, R. M. et al. Domestication and large animal interactions: skeletal trauma in northern Vietnam during the hunter-gatherer Da But period. *PLoS ONE* **14**, e0218777 (2019).

13. Vlok, M., Paz, V., Crozier, R. & Oxenham, M. A new application of the bioarchaeology of care approach: a case study from the metal period, the Philippines. *Int. J. Osteoarchaeol.* **27**, 662–671 (2017).

14. Rouhani, A., Elmi, A., Aghdam, H. A., Panahi, F. & Ghafari, Y. D. The role of fibular fixation in the treatment of tibia diaphysis distal third fractures. *Orthop. Traumatol. Surg. Res.* **98**, 868–872 (2012).

15. Bonnevialle, P. et al. Distal leg fractures: how critical is the fibular fracture and its fixation? *Orthop. Traumatol. Surg. Res.* **96**, 667–673 (2010).

16. Donnally, C. et al. Orthopedic injuries associated with jet-skis (personal watercrafts): a review of 127 inpatients. *Orthop. Traumatol. Surg. Res.* **104**, 267–271 (2018).

17. Pennoyer, G. P. Traumatic amputation of the thigh, complicated by both tetanus and gas gangrene with recovery. *JAMA* **95**, 342–343 (1930).

18. Wen, A. P. Y. et al. Successful ankle replantation in two cases with different presentations. *Arch. Plast. Surg.* **47**, 182–186 (2020).

19. Pannier, S. Congenital pseudarthrosis of the tibia. *Orthop. Traumatol. Surg. Res.* **97**, 750–761 (2011).

20. Abbas, Z. G. & Archibald, L. K. Epidemiology of the diabetic foot in Africa. *Med. Sci. Monit.* **11**, RA262–RA270 (2005).

21. Eisenberg, K. A. & Vuillermin, C. B. Management of congenital pseudoarthrosis of the tibia and fibula. *Curr. Rev. Musculoskelet. Med.* **12**, 356–368 (2019).

22. Anderson, E., Peluso, S., Lettice, L. A. & Hill, R. E. Human limb abnormalities caused by disruption of hedgehog signaling. *Trends Genet.* **28**, 364–373 (2012).

23. Tadmor, O., Kreisbe, G., Achiront, R., Porat, S. & Yagel, S. Limb amputation in amniotic band syndrome: serial ultrasonographic and Doppler observations. *Ultrasound Obstet. Gynecol.* **10**, 312–315 (1997).

24. Baker, C. J. & Rudolph, A. J. Congenital ring constrictions and intrauterine amputations. *Am. J. Dis. Child.* **121**, 393–400 (1971).

25. Fisher, R. & Cremin, B. Limb defects in the amniotic band syndrome. *Pediatr. Radiol.* **5**, 24–29 (1976).

26. Barber, C. G. Immediate and eventual features of healing in amputated bones. *Ann. Surg.* **90**, 985–992 (1929).

27. Burns, K. R. *Forensic Anthropology Training Manual* (Routledge, 2015).

## Acknowledgements

We thank H. Rice for his collaboration on producing the CT scan data.

## Author information

Author notes

1. These authors contributed equally: Melandri Vlok, Tim Maloney, India Ella Dilkes-Hall, Adhi Agus Oktaviana, Andika Arief Drajat Priyatno, Maxime Aubert

2. Deceased: Pindi Setiawan

### Authors and Affiliations

1. Sydney Southeast Asian Centre, University of Sydney, Sydney, New South Wales, Australia

   Melandri Vlok

2. Griffith Centre for Social and Cultural Research, Griffith University, Gold Coast, Queensland, Australia

   Tim Maloney & Maxime Aubert

3. Australian Research Centre for Human Evolution, Griffith University, Nathan, Queensland, Australia

   Tim Maloney, Adam Brumm & Maxime Aubert

4. Research into Deer Genetics and Environment, RIDGE Group Inc, Ascot, Western Australia, Australia

Tim Maloney

5. Archaeology, School of Social Sciences, University of Western Australia, Crawley, Western Australia, Australia

India Ella Dilkes-Hall

6. School of Humanities, Languages and Social Science, Griffith University, Gold Coast, Queensland, Australia

Adhi Agus Oktaviana & Shinatria Adhityatama

7. BRIN, OR Arkeologi, Bahasa dan Sastra, Pusat Riset Arkeometri, Jakarta, Indonesia

Adhi Agus Oktaviana, Marlon Ririmasse & I. Made Geria

8. Faculty of Art and Design, Bandung Institute of Technology, Bandung, Indonesia

Pindi Setiawan

9. Balai Pelestarian Cagar Budaya Kalimantan Timur, Samarinda, Indonesia

Andika Arief Drajat Priyatno, Muslimin A. R. Effendy, Budy Istiawan & Falentinus Triwijaya Atmoko

10. Archaeology, College of Humanities, Arts and Social Sciences, Flinders University, Bedford Park, South Australia, Australia

Ian Moffat

11. Geoarchaeology and Archaeometry Research Group (GARG), Southern Cross University, Lismore, New South Wales, Australia

Renaud Joannes-Boyau & Maxime Aubert

12. Palaeo-Research Institute, University of Johannesburg, Johannesburg, South Africa

Renaud Joannes-Boyau

### Contributions

M.V., T.M., I.E.D.-H. and A.A.D.P. conceived and wrote the paper. T.M., I.E.D.-H. and A.A.D.P carried out the excavation of the site and burial. M.A. and A.B. conceived the study and contributed to the paper. Site access, project coordination and field logistics were facilitated by P.S., M.R., A.A.O., F.T.A., I.M.G., M.A.R.E., B.I. and S.A. M.V. conducted the osteological analyses. I.M. conducted the geophysical survey. R.J.-B. conducted the US–ESR dating analyses and the Bayesian modelling. All of the authors contributed to editing the paper.

### Corresponding authors

Correspondence to Melandri Vlok or Tim Maloney.

## Ethics declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Extended data figures and tables

**Extended Data Table 1 Differential Diagnosis of Tebo 1**

# Supplementary information

## Reporting Summary

# Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

# About this article

## Cite this article

Vlok, M., Maloney, T., Dilkes-Hall, I.E. *et al.* Reply to: Common orthopaedic trauma may explain 31,000-year-old remains. *Nature* **615**, E15–E18 (2023). https://doi.org/10.1038/s41586-023-05757-7

- Published: 15 March 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05757-7

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

## Surgical amputation of a limb 31,000 years ago in Borneo

- Tim Ryan Maloney
- India Ella Dilkes-Hall
- Maxime Aubert

Nature Article Open Access 07 Sept 2022

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05757-7

# Amendments & Corrections

- **Author Correction: FinnGen provides genetic insights from a well-phenotyped isolated population** [24 Feb 2023]
    -
- **Author Correction: Stigma receptors control intraspecies and interspecies barriers in Brassicaceae** [02 Mar 2023]
    -
- **Publisher Correction: Heterogeneous melting near the Thwaites Glacier grounding line** [24 Feb 2023]
    -
- **Retraction Note: Site-specific group selection drives locally adapted group compositions** [06 Mar 2023]
    -

- Author Correction
- Open Access
- Published: 24 February 2023


# Author Correction: FinnGen provides genetic insights from a well-phenotyped isolated population


- Mitja I. Kurki,
- Juha Karjalainen,
- Priit Palta,
- Timo P. Sipilä,
- Kati Kristiansson,
- Kati M. Donner,
- Mary P. Reeve,
- Hannele Laivuori,
- Mervi Aavikko,
- Mari A. Kaunisto,
- Anu Loukola,
- Elisa Lahtela,
- Hannele Mattsson,
- Päivi Laiho,
- Pietro Della Briotta Parolo,
- Arto A. Lehisto,
- Masahiro Kanai,
- Nina Mars,
- Joel Rämö,
- Tuomo Kiiskinen,
- Henrike O. Heyne,

- Kumar Veerapen,
- Sina Rüeger,
- Susanna Lemmelä,
- Wei Zhou,
- Sanni Ruotsalainen,
- Kalle Pärn,
- Tero Hiekkalinna,
- Sami Koskelainen,
- Teemu Paajanen,
- Vincent Llorens,
- Javier Gracia-Tabuenca,
- Harri Siirtola,
- Kadri Reis,
- Abdelrahman G. Elnahas,
- Benjamin Sun,
- Christopher N. Foley,
- Katriina Aalto-Setälä,
- Kaur Alasoo,
- Mikko Arvas,
- Kirsi Auro,
- Shameek Biswas,
- Argyro Bizaki-Vallaskangas,
- Olli Carpen,
- Chia-Yen Chen,
- Oluwaseun A. Dada,
- Zhihao Ding,
- Margaret G. Ehm,
- Kari Eklund,
- Martti Färkkilä,
- Hilary Finucane,
- Andrea Ganna,
- Awaisa Ghazal,
- Robert R. Graham,
- Eric M. Green,
- Antti Hakanen,
- Marco Hautalahti,
- Åsa K. Hedman,

- Mikko Hiltunen,
- Reetta Hinttala,
- Iiris Hovatta,
- Xinli Hu,
- Adriana Huertas-Vazquez,
- Laura Huilaja,
- Julie Hunkapiller,
- Howard Jacob,
- Jan-Nygaard Jensen,
- Heikki Joensuu,
- Sally John,
- Valtteri Julkunen,
- Marc Jung,
- Juhani Junttila,
- Kai Kaarniranta,
- Mika Kähönen,
- Risto Kajanne,
- Lila Kallio,
- Reetta Kälviäinen,
- Jaakko Kaprio,
- FinnGen,
- Nurlan Kerimov,
- Johannes Kettunen,
- Elina Kilpeläinen,
- Terhi Kilpi,
- Katherine Klinger,
- Veli-Matti Kosma,
- Teijo Kuopio,
- Venla Kurra,
- Triin Laisk,
- Jari Laukkanen,
- Nathan Lawless,
- Aoxing Liu,
- Simonne Longerich,
- Reedik Mägi,
- Johanna Mäkelä,
- Antti Mäkitie,

- Anders Malarstig,
- Arto Mannermaa,
- Joseph Maranville,
- Athena Matakidou,
- Tuomo Meretoja,
- Sahar V. Mozaffari,
- Mari E. K. Niemi,
- Marianna Niemi,
- Teemu Niiranen,
- Christopher J. O´Donnell,
- Ma´en Obeidat,
- George Okafo,
- Hanna M. Ollila,
- Antti Palomäki,
- Tuula Palotie,
- Jukka Partanen,
- Dirk S. Paul,
- Margit Pelkonen,
- Rion K. Pendergrass,
- Slavé Petrovski,
- Anne Pitkäranta,
- Adam Platt,
- David Pulford,
- Eero Punkka,
- Pirkko Pussinen,
- Neha Raghavan,
- Fedik Rahimov,
- Deepak Rajpal,
- Nicole A. Renaud,
- Bridget Riley-Gillis,
- Rodosthenis Rodosthenous,
- Elmo Saarentaus,
- Aino Salminen,
- Eveliina Salminen,
- Veikko Salomaa,
- Johanna Schleutker,
- Raisa Serpi,

- Huei-yi Shen,
- Richard Siegel,
- Kaisa Silander,
- Sanna Siltanen,
- Sirpa Soini,
- Hilkka Soininen,
- Jae Hoon Sul,
- Ioanna Tachmazidou,
- Kaisa Tasanen,
- Pentti Tienari,
- Sanna Toppila-Salmi,
- Taru Tukiainen,
- Tiinamaija Tuomi,
- Joni A. Turunen,
- Jacob C. Ulirsch,
- Felix Vaura,
- Petri Virolainen,
- Jeffrey Waring,
- Dawn Waterworth,
- Robert Yang,
- Mari Nelis,
- Anu Reigo,
- Andres Metspalu,
- Lili Milani,
- Tõnu Esko,
- Caroline Fox,
- Aki S. Havulinna,
- Markus Perola,
- Samuli Ripatti,
- Anu Jalanko,
- Tarja Laitinen,
- Tomi P. Mäkelä,
- Robert Plenge,
- Mark McCarthy,
- Heiko Runz,
- Mark J. Daly &
- …

- Aarno Palotie

Show authors

*Nature* **volume 615**, page E19 (2023)

- 472 Accesses
- 1 Citations
- 1 Altmetric
- Metrics details

The Original Article was published on 18 January 2023

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05473-8
Published online 18 January 2023

In the version of this article initially published, author Marianna Niemi, now affiliated with the TAUCHI Research Center and the Faculty of Medicine and Health Technology in Tampere University, was mistakenly affiliated with TAUCHI Research Center within the Faculty of Information Technology and Communication Sciences department. The error has been corrected in the HTML and PDF versions of the article.

# Author information

Author notes

1. These authors jointly supervised this work: Mark J. Daly, Aarno Palotie

## Authors and Affiliations

1. Institute for Molecular Medicine Finland (FIMM), Helsinki Institute of Life Science (HiLIFE), University of Helsinki, Helsinki, Finland

Mitja I. Kurki, Juha Karjalainen, Priit Palta, Timo P. Sipilä, Kati M. Donner, Mary P. Reeve, Hannele Laivuori, Mervi Aavikko, Mari A. Kaunisto, Elisa Lahtela, Pietro Della Briotta Parolo, Arto A. Lehisto, Masahiro Kanai, Nina Mars, Joel Rämö, Tuomo Kiiskinen, Henrike O. Heyne, Kumar Veerapen, Sina Rüeger, Susanna Lemmelä, Sanni Ruotsalainen, Kalle Pärn, Vincent Llorens, Oluwaseun A. Dada, Andrea Ganna, Awaisa Ghazal, Risto Kajanne, Jaakko Kaprio, Elina Kilpeläinen, Aoxing Liu, Mari E. K. Niemi, Hanna M. Ollila, Rodosthenis Rodosthenous, Elmo Saarentaus, Huei-yi Shen, Taru Tukiainen, Tiinamaija Tuomi, Aki S. Havulinna, Samuli Ripatti, Anu Jalanko, Mark J. Daly & Aarno Palotie

2. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

Mitja I. Kurki, Juha Karjalainen, Masahiro Kanai, Henrike O. Heyne, Kumar Veerapen, Wei Zhou, Hilary Finucane, Andrea Ganna, Jacob C. Ulirsch, Mark J. Daly & Aarno Palotie

3. Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA

Mitja I. Kurki, Juha Karjalainen, Masahiro Kanai, Henrike O. Heyne, Kumar Veerapen, Wei Zhou, Hilary Finucane, Andrea Ganna, Jacob C. Ulirsch, Mark J. Daly & Aarno Palotie

4. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

Mitja I. Kurki, Juha Karjalainen, Masahiro Kanai, Kumar Veerapen, Wei Zhou, Hilary Finucane, Andrea Ganna, Mark J. Daly & Aarno Palotie

5. Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

Priit Palta, Kadri Reis, Abdelrahman G. Elnahas, Triin Laisk, Reedik Mägi, Anu Reigo, Andres Metspalu, Lili Milani & Tõnu Esko

6. Finnish Institute for Health and Welfare (THL), Helsinki, Finland

Kati Kristiansson, Hannele Mattsson, Päivi Laiho, Susanna Lemmelä, Tero Hiekkalinna, Sami Koskelainen, Teemu Paajanen, Johannes Kettunen, Terhi Kilpi, Teemu Niiranen, Veikko Salomaa, Kaisa Silander, Sirpa Soini, Felix Vaura, Aki S. Havulinna & Markus Perola

7. Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Hannele Laivuori

8. Department of Obstetrics and Gynecology, Tampere University Hospital, Tampere, Finland

Hannele Laivuori

9. Faculty of Medicine and Health Technology, Center for Child, Adolescent and Maternal Health, University of Tampere, Tampere, Finland

Hannele Laivuori

10. Helsinki Biobank, University of Helsinki and Hospital District of Helsinki and Uusimaa, Helsinki, Finland

Anu Loukola, Olli Carpen & Eero Punkka

11. Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

Masahiro Kanai

12. Digital Health Center, Hasso Plattner Institute for Digital Engineering, University of Potsdam Potsdam, Potsdam, Germany

Henrike O. Heyne

13. Hasso Plattner Institute for Digital Health at Mount Sinai, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Henrike O. Heyne

14. TAUCHI Research Center, Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland

Javier Gracia-Tabuenca & Harri Siirtola

15. Translational Biology, Research and Development, Biogen, Cambridge, MA, USA

Benjamin Sun

16. BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Benjamin Sun

17. Optima Partners, Edinburgh, UK

Christopher N. Foley

18. MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Cambridge, UK

Christopher N. Foley

19. Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

Katriina Aalto-Setälä, Mika Kähönen & Marianna Niemi

20. Institute of Computer Science, University of Tartu, Tartu, Estonia

Kaur Alasoo & Nurlan Kerimov

21. Finnish Red Cross Blood Service, Helsinki, Finland

Mikko Arvas & Jukka Partanen

22. GlaxoSmithKline, Espoo, Finland

Kirsi Auro

23. Bristol Myers Squibb, New York, NY, USA

Shameek Biswas, Joseph Maranville & Robert Plenge

24. Tampere University Hospital and Tampere University, Tampere, Finland

Argyro Bizaki-Vallaskangas

25. Biogen, Cambridge, MA, USA

Chia-Yen Chen, Sally John & Heiko Runz

26. Boehringer Ingelheim, Ingelheim am Rhein, Germany

Zhihao Ding, Jan-Nygaard Jensen, Marc Jung, Nathan Lawless & George Okafo

27. GlaxoSmithKline, Collegeville, PA, USA

Margaret G. Ehm

28. Division of Rheumatology, Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland

Kari Eklund

29. Orton Orthopedic Hospital, Helsinki, Finland

Kari Eklund

30. Abdominal Center, Helsinki University Hospital, Helsinki University, Helsinki, Finland

Martti Färkkilä

31. Maze Therapeutics, South San Francisco, CA, USA

Robert R. Graham, Eric M. Green & Sahar V. Mozaffari

32. Auria Biobank, University of Turku and Turku University Hospital, Turku, Finland

Antti Hakanen, Lila Kallio, Johanna Schleutker & Petri Virolainen

33. FINBB, Finnish Biobank Cooperative, Helsinki, Finland

Marco Hautalahti

34. Pfizer, New York, NY, USA

Åsa K. Hedman & Xinli Hu

35. Department of Medicine, Karolinska Institute, Solna, Sweden

Åsa K. Hedman

36. Clinical Biobank Tampere, Tampere University and Tampere University Hospital, Tampere, Finland

Mikko Hiltunen

37. Medical Research Center Oulu and PEDEGO Research Unit, University of Oulu, Oulu, Finland

Reetta Hinttala

38. Biocenter Oulu, University of Oulu, Oulu, Finland

Reetta Hinttala & Johannes Kettunen

39. Oulu University Hospital, Oulu, Finland

Reetta Hinttala

40. Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

Iiris Hovatta

41. SleepWell Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland

Iiris Hovatta

42. Merck & Co, Kenilworth, NJ, USA

Adriana Huertas-Vazquez, Simonne Longerich, Neha Raghavan, Jae Hoon Sul & Caroline Fox

43. PEDEGO Research Unit, University of Oulu, Oulu, Finland

Laura Huilaja & Kaisa Tasanen

44. Department of Dermatology and Medical Research Center Oulu, Oulu University Hospital, Oulu, Finland

Laura Huilaja & Kaisa Tasanen

45. Genentech, San Francisco, CA, USA

Julie Hunkapiller, Rion K. Pendergrass & Mark McCarthy

46. AbbVie, Chicago, IL, USA

Howard Jacob, Fedik Rahimov, Bridget Riley-Gillis & Jeffrey Waring

47. Helsinki University Hospital and University of Helsinki, Helsinki, Finland

Heikki Joensuu, Tuomo Meretoja & Joni A. Turunen

48. Neuro Center, Neurology, Kuopio University Hospital, Kuopio, Finland

Valtteri Julkunen

49. Institute of Clinical Medicine–Neurology, University of Eastern Finland, Kuopio, Finland

Valtteri Julkunen

50. Northern Finland Biobank Borealis, University of Oulu, Northern Ostrobothnia Hospital District, Oulu, Finland

Juhani Junttila & Raisa Serpi

51. Department of Ophthalmology, Kuopio University Hospital, Kuopio, Finland

Kai Kaarniranta

52. Department of Ophthalmology, Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland

Kai Kaarniranta

53. Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland

Mika Kähönen

54. Epilepsy Center, Kuopio University Hospital, Kuopio, Finland

Reetta Kälviäinen

55. Department of Neurology, University of Eastern Finland, Kuopio, Finland

Reetta Kälviäinen

56. Department of Public Health, University of Helsinki, Helsinki, Finland

Jaakko Kaprio

57. Computational Medicine, Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland

Johannes Kettunen

58. Translational Sciences, Sanofi R&D, Framingham, MA, USA

Katherine Klinger & Deepak Rajpal

59. Biobank of Eastern Finland, University of Eastern Finland, Kuopio, Finland

Veli-Matti Kosma & Arto Mannermaa

60. Kuopio University Hospital, Kuopio, Finland

Veli-Matti Kosma & Arto Mannermaa

61. Central Finland Biobank, Central Finland Health Care District, Jyväskylä, Finland

Teijo Kuopio & Jari Laukkanen

62. Department of Clinical Genetics, Tampere University Hospital, Tampere, Finland

Venla Kurra

63. Department of Clinical Genetics, Faculty of Medicine and Health Technology, Tampere, Finland

Venla Kurra

64. Department of Medicine, Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland

Jari Laukkanen

65. FINBB, Finnish Biobank Cooperative, Turku, Finland

Johanna Mäkelä

66. Department of Otorhinolaryngology–Head and Neck Surgery, University of Helsinki, Helsinki, Finland

Antti Mäkitie

67. Helsinki University Hospital, Helsinki, Finland

Antti Mäkitie & Eveliina Salminen

68. Pfizer, Cambridge, MA, USA

Anders Malarstig

69. Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Solna, Sweden

Anders Malarstig

70. Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK

Athena Matakidou, Dirk S. Paul, Slavé Petrovski & Ioanna Tachmazidou

71. TAUCHI Research Center & Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

Marianna Niemi

72. Turku University Hospital and University of Turku, Turku, Finland

Teemu Niiranen & Antti Palomäki

73. Novartis Institutes for BioMedical Research, Cambridge, MA, USA

Christopher J. O´Donnell, Ma´en Obeidat & Nicole A. Renaud

74. Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA

Hanna M. Ollila

75. Department of Oral and Maxillofacial Diseases, Helsinki University Hospital, Helsinki, Finland

Tuula Palotie

76. Department of Oral and Maxillofacial Diseases, University of Helsinki, Helsinki, Finland

Tuula Palotie, Pirkko Pussinen & Aino Salminen

77. Finnish Hematological Biobank, Helsinki, Finland

Jukka Partanen

78. Department of Pulmonary Diseases, Kuopio University Hospital, Kuopio, Finland

Margit Pelkonen

79. Department of Otorhinolaryngology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

Anne Pitkäranta

80. Translational Science and Experimental Medicine, Research and Early Development, Respiratory and Immunology, BioPharmaceuticals

R&D, AstraZeneca, Cambridge, UK

Adam Platt

81. GlaxoSmithKline, Stevenage, UK

David Pulford

82. Department of Clinical Genetics, HUSLAB, HUS Diagnostic Center, University of Helsinki, Helsinki, Finland

Eveliina Salminen

83. Novartis Institutes for BioMedical Research, Basel, Switzerland

Richard Siegel

84. Finnish Clinical Biobank Tampere, Tampere University and Tampere University Hospital, Tampere, Finland

Sanna Siltanen & Tarja Laitinen

85. Department of Neurology, Institute of Clinical Medicine, University of Eastern Finland, Kuopio, Finland

Hilkka Soininen

86. Department of Neurology, Helsinki University Hospital, Helsinki, Finland

Pentti Tienari

87. Translational Immunology, Research Programs Unit, University of Helsinki, Helsinki, Finland

Pentti Tienari

88. Department of Allergy, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

Sanna Toppila-Salmi

89. Abdominal Center, Endocrinology, Helsinki University Hospital, Helsinki, Finland

Tiinamaija Tuomi

90. Folkhalsan Research Center, Helsinki, Finland

Tiinamaija Tuomi

91. Research Program of Clinical and Molecular Metabolism, University of Helsinki, Helsinki, Finland

Tiinamaija Tuomi

92. Eye Genetics Group, Folkhälsan Research Center, Helsinki, Finland

Joni A. Turunen

93. University of Turku, Turku, Finland

Felix Vaura

94. Janssen Research & Development, Spring House, PA, USA

Dawn Waterworth

95. Janssen Biotech, Beerse, Belgium

Robert Yang

96. Genomics Core Facility, Institute of Genomics, University of Tartu, Tartu, Estonia

Mari Nelis

97. Helsinki Institute of Life Science (HiLIFE), University of Helsinki, Helsinki, Finland

Tomi P. Mäkelä

## Consortia

## FinnGen

- Mitja I. Kurki
- , Juha Karjalainen
- , Priit Palta
- , Timo P. Sipilä
- , Kati Kristiansson
- , Kati M. Donner
- , Mary P. Reeve
- , Hannele Laivuori
- , Mervi Aavikko
- , Mari A. Kaunisto
- , Anu Loukola
- , Elisa Lahtela
- , Hannele Mattsson
- , Päivi Laiho
- , Pietro Della Briotta Parolo
- , Arto A. Lehisto
- , Masahiro Kanai
- , Nina Mars
- , Joel Rämö
- , Tuomo Kiiskinen
- , Henrike O. Heyne
- , Kumar Veerapen
- , Sina Rüeger
- , Susanna Lemmelä
- , Wei Zhou
- , Sanni Ruotsalainen
- , Kalle Pärn
- , Tero Hiekkalinna
- , Sami Koskelainen
- , Teemu Paajanen
- , Vincent Llorens

- , Javier Gracia-Tabuenca
- , Harri Siirtola
- , Benjamin Sun
- , Katriina Aalto-Setälä
- , Mikko Arvas
- , Kirsi Auro
- , Shameek Biswas
- , Argyro Bizaki-Vallaskangas
- , Olli Carpen
- , Chia-Yen Chen
- , Oluwaseun A. Dada
- , Zhihao Ding
- , Margaret G. Ehm
- , Kari Eklund
- , Martti Färkkilä
- , Hilary Finucane
- , Andrea Ganna
- , Awaisa Ghazal
- , Robert R. Graham
- , Eric M. Green
- , Antti Hakanen
- , Marco Hautalahti
- , Åsa K. Hedman
- , Mikko Hiltunen
- , Reetta Hinttala
- , Iiris Hovatta
- , Xinli Hu
- , Adriana Huertas-Vazquez
- , Laura Huilaja
- , Julie Hunkapiller
- , Howard Jacob
- , Jan-Nygaard Jensen
- , Heikki Joensuu
- , Sally John
- , Valtteri Julkunen
- , Marc Jung
- , Juhani Junttila

- , Kai Kaarniranta
- , Mika Kähönen
- , Risto Kajanne
- , Lila Kallio
- , Reetta Kälviäinen
- , Jaakko Kaprio
- , Nurlan Kerimov
- , Johannes Kettunen
- , Elina Kilpeläinen
- , Terhi Kilpi
- , Katherine Klinger
- , Veli-Matti Kosma
- , Teijo Kuopio
- , Venla Kurra
- , Jari Laukkanen
- , Nathan Lawless
- , Aoxing Liu
- , Simonne Longerich
- , Johanna Mäkelä
- , Antti Mäkitie
- , Anders Malarstig
- , Arto Mannermaa
- , Joseph Maranville
- , Athena Matakidou
- , Tuomo Meretoja
- , Sahar V. Mozaffari
- , Mari E. K. Niemi
- , Marianna Niemi
- , Teemu Niiranen
- , Christopher J. O´Donnell
- , Ma´en Obeidat
- , George Okafo
- , Hanna M. Ollila
- , Antti Palomäki
- , Tuula Palotie
- , Jukka Partanen
- , Dirk S. Paul

- , Margit Pelkonen
- , Rion K. Pendergrass
- , Slavé Petrovski
- , Anne Pitkäranta
- , Adam Platt
- , David Pulford
- , Eero Punkka
- , Pirkko Pussinen
- , Neha Raghavan
- , Fedik Rahimov
- , Deepak Rajpal
- , Nicole A. Renaud
- , Bridget Riley-Gillis
- , Rodosthenis Rodosthenous
- , Elmo Saarentaus
- , Aino Salminen
- , Eveliina Salminen
- , Veikko Salomaa
- , Johanna Schleutker
- , Raisa Serpi
- , Huei-yi Shen
- , Richard Siegel
- , Kaisa Silander
- , Sanna Siltanen
- , Sirpa Soini
- , Hilkka Soininen
- , Jae Hoon Sul
- , Ioanna Tachmazidou
- , Kaisa Tasanen
- , Pentti Tienari
- , Sanna Toppila-Salmi
- , Taru Tukiainen
- , Tiinamaija Tuomi
- , Joni A. Turunen
- , Jacob C. Ulirsch
- , Felix Vaura
- , Petri Virolainen

- , Jeffrey Waring
- , Dawn Waterworth
- , Robert Yang
- , Caroline Fox
- , Aki S. Havulinna
- , Markus Perola
- , Samuli Ripatti
- , Anu Jalanko
- , Tarja Laitinen
- , Tomi P. Mäkelä
- , Robert Plenge
- , Mark McCarthy
- , Heiko Runz
- , Mark J. Daly
-  & Aarno Palotie

### Corresponding author

Correspondence to Aarno Palotie.

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

# About this article

## Cite this article

Kurki, M.I., Karjalainen, J., Palta, P. *et al.* Author Correction: FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **615**, E19 (2023). https://doi.org/10.1038/s41586-023-05837-8


- Published: 24 February 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05837-8


## Share this article


Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05837-8

- Author Correction
- Open Access
- Published: 02 March 2023

# Author Correction: Stigma receptors control intraspecies and interspecies barriers in Brassicaceae


- Jiabao Huang,
- Lin Yang,
- Liu Yang,
- Xiaoyu Wu,
- Xiaoshuang Cui,
- Lili Zhang,
- Jiyun Hui,
- Yumei Zhao,
- Hongmin Yang,
- Shangjia Liu,
- Quanling Xu,
- Maoxuan Pang,
- Xinping Guo,
- Yunyun Cao,
- Yu Chen,
- Xinru Ren,
- Jinzhi Lv,
- Jianqiang Yu,
- Junyi Ding,
- Gang Xu,
- Nian Wang,

- Xiaochun Wei,
- Qinghui Lin,
- Yuxiang Yuan,
- Xiaowei Zhang,
- Chaozhi Ma,
- Cheng Dai,
- Pengwei Wang,
- Yongchao Wang,
- Fei Cheng,
- Weiqing Zeng,
- Ravishankar Palanivelu,
- Hen-Ming Wu,
- Xiansheng Zhang,
- Alice Y. Cheung &
- …
- Qiaohong Duan

Show authors

*Nature* **volume 615**, page E20 (2023)

- 407 Accesses
- 1 Altmetric
- Metrics details

The Original Article was published on 25 January 2023

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05640-x
Published online 25 January 2023

In the version of this article initially published, the affiliation listed for Weiqing Zeng was incorrect (Trait Discovery, Corteva Agriscience, Johnston, IA, USA); it should have read International Flavors &

Fragrances, Wilmington, DE, USA. The affiliation has been corrected in the HTML and PDF versions of the article.

# Author information

Author notes

1. These authors contributed equally: Jiabao Huang, Lin Yang


## Authors and Affiliations

1. State Key Laboratory of Crop Biology, Shandong Agricultural University, Tai'an, China

   Jiabao Huang, Lin Yang, Liu Yang, Xiaoyu Wu, Xiaoshuang Cui, Lili Zhang, Jiyun Hui, Yumei Zhao, Hongmin Yang, Shangjia Liu, Quanling Xu, Maoxuan Pang, Xinping Guo, Yunyun Cao, Yu Chen, Xinru Ren, Jinzhi Lv, Jianqiang Yu, Junyi Ding, Gang Xu, Nian Wang, Xiansheng Zhang & Qiaohong Duan

2. College of Horticulture Science and Engineering, Shandong Agricultural University, Tai'an, China

   Jiabao Huang, Lin Yang, Liu Yang, Xiaoyu Wu, Xiaoshuang Cui, Lili Zhang, Jiyun Hui, Yumei Zhao, Hongmin Yang, Shangjia Liu, Quanling Xu, Maoxuan Pang, Xinping Guo, Yunyun Cao, Yu Chen, Xinru Ren, Jinzhi Lv, Jianqiang Yu & Qiaohong Duan

3. Institute of Horticulture, Henan Academy of Agricultural Sciences, Zhengzhou, China

   Xiaochun Wei, Yuxiang Yuan & Xiaowei Zhang

4. Computer Network Information Centre, Chinese Academy of Sciences, Beijing, China

   Qinghui Lin

5. National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

   Chaozhi Ma, Cheng Dai & Pengwei Wang

6. Shandong Yiyi Agricultural Science and Technology Co., Ltd, Tai'an, China

   Yongchao Wang

7. College of Horticulture, Qingdao Agricultural University, Qingdao, China

   Fei Cheng

8. International Flavors & Fragrances, Wilmington, DE, USA

   Weiqing Zeng

9. School of Plant Sciences, University of Arizona, Tucson, AZ, USA

   Ravishankar Palanivelu

10. Department of Biochemistry and Molecular Biology, Molecular Cell Biology and Plant Biology Programs, University of Massachusetts, Amherst, MA, USA

   Hen-Ming Wu & Alice Y. Cheung

### Corresponding authors

Correspondence to Alice Y. Cheung or Qiaohong Duan.

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long

as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

# About this article

## Cite this article

Huang, J., Yang, L., Yang, L. *et al.* Author Correction: Stigma receptors control intraspecies and interspecies barriers in Brassicaceae. *Nature* **615**, E20 (2023). https://doi.org/10.1038/s41586-023-05816-z

- Published: 02 March 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05816-z

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05816-z

- Publisher Correction
- Open Access
- Published: 24 February 2023


# Publisher Correction: Heterogeneous melting near the Thwaites Glacier grounding line


- B. E. Schmidt ,
- P. Washam,
- P. E. D. Davis,
- K. W. Nicholls,
- D. M. Holland,
- J. D. Lawrence,
- K. L. Riverman,
- J. A. Smith,
- A. Spears,
- D. J. G. Dichek,
- A. D. Mullen,
- E. Clyne,
- B. Yeager,
- P. Anker,
- M. R. Meister,
- B. C. Hurwitz,
- E. S. Quartini,
- F. E. Bryson,
- A. Basinski-Ferris,
- C. Thomas,
- J. Wake,
- D. G. Vaughan,
- S. Anandakrishnan,
- E. Rignot,

- J. Paden &
- …
- K. Makinson

Show authors

*Nature* **volume 615**, page E21 (2023)

- 509 Accesses
- 11 Altmetric
- Metrics details

The Original Article was published on 15 February 2023

Correction to: *Nature* https://doi.org/10.1038/s41586-022-05691-0
Published online 15 February 2023

In the version of this article originally published, units on the colour bar in Fig. 4a, now reading "6, 3, 0," were initially jumbled, reading "0, 36". In addition, the scale was inverted. The units and scale have been corrected in the HTML and PDF versions of the article.

# Author information

## Authors and Affiliations

1. Department of Astronomy, Cornell University, Ithaca, NY, USA

   B. E. Schmidt, P. Washam, D. J. G. Dichek, A. D. Mullen, M. R. Meister, E. S. Quartini & F. E. Bryson

2. Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, USA

B. E. Schmidt, P. Washam, D. J. G. Dichek, A. D. Mullen, M. R. Meister, E. S. Quartini & F. E. Bryson

3. British Antarctic Survey, Cambridge, UK

P. E. D. Davis, K. W. Nicholls, J. A. Smith, P. Anker, C. Thomas, J. Wake, D. G. Vaughan & K. Makinson

4. Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

D. M. Holland & A. Basinski-Ferris

5. Center for Global Sea Level Change, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

D. M. Holland & B. Yeager

6. School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA

J. D. Lawrence, A. Spears, B. C. Hurwitz & F. E. Bryson

7. Department of Environmental Studies, University of Portland, Portland, OR, USA

K. L. Riverman

8. Department of Geosciences, Pennsylvania State University, State College, PA, USA

E. Clyne & S. Anandakrishnan

9. Environmental Studies, Lewis & Clark College, Portland, OR, USA

E. Clyne

10. Department of Earth System Science, University of California, Irvine, Irvine, CA, USA

E. Rignot

11. Center for Remote Sensing and Integrated Systems, University of Kansas, Lawrence, KS, USA

J. Paden

### Corresponding author

Correspondence to B. E. Schmidt.

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

## About this article

### Cite this article

Schmidt, B.E., Washam, P., Davis, P.E.D. *et al.* Publisher Correction: Heterogeneous melting near the Thwaites Glacier grounding line. *Nature* **615**, E21 (2023). https://doi.org/10.1038/s41586-023-05861-8

- Published: 24 February 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05861-8

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05861-8

- Retraction Note
- Published: 06 March 2023

# Retraction Note: Site-specific group selection drives locally adapted group compositions


- Jonathan N. Pruitt &
- Charles J. Goodnight


*Nature* **volume 615**, page 548 (2023)


- 9682 Accesses

- 45 Altmetric

- Metrics details


The Original Article was published on 01 October 2014

Retraction of: *Nature* https://doi.org/10.1038/nature13811 Published online 1 October 2014

The Editors have retracted this article. Concerns were raised regarding potential anomalies in the census and experimental data on the aggressiveness and docility of the spiders and the experimental data on prey availability. Post-publication review concluded that the data are not reliable. The Editors therefore no longer have confidence in the results and conclusions of this article. Jonathan N. Pruitt disagrees with this retraction. Charles J. Goodnight is deceased.

# Author information

## Authors and Affiliations

1. Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

   Jonathan N. Pruitt

2. Department of Biology, University of Vermont, Burlington, Vermont, USA

   Charles J. Goodnight

## Corresponding author

Correspondence to Jonathan N. Pruitt.

# About this article

## Cite this article

Pruitt, J.N., Goodnight, C.J. Retraction Note: Site-specific group selection drives locally adapted group compositions. *Nature* **615**, 548 (2023). https://doi.org/10.1038/s41586-023-05882-3

- Published: 06 March 2023
- Issue Date: 16 March 2023
- DOI: https://doi.org/10.1038/s41586-023-05882-3

## Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05882-3